omega

The International Journal of Management Science

Volume 6 1978

Author and Title Index

● **Chief Editor**
Professor SAMUEL EILON,
DSc(Eng.), PhD
Department of Management Science, Imperial College, London, SW7 2BX

● **Associate Editors**
RJ BETTS and **RB FLAVELL**
Department of Management Science, Imperial College, London, SW7 2BX

Editorial Advisory Board

Professor RJ BALL
London Graduate School of Business Studies, Sussex Place, Regent's Park, London NW1 4SA

Professor BEN BERNHOLTZ
Department of Industrial Engineering, University of Toronto, Toronto 5

Professor C WEST CHURCHMAN
Space Sciences Laboratory, University of California, Berkeley, CA 94720

Professor WW COOPER
Graduate School of Business Administration, Harvard University, Boston, MA 02163

Professor BELA GOLD
Department of Economics, Case Western Reserve University, Cleveland, OH 44106

Professor CH GUDNASON
Danmarks Tekniske Højskole, Kobenhavn

Professor DOUGLAS C HAGUE
Manchester Business School, Booth Street West, Manchester M15 6PB

DAVID B HERTZ
McKinsey and Company Inc., 245 Park Avenue, New York, NY 10017

Professor KNUT HOLT
The Norwegian Institute of Technology, 7034 Trondheim

Professor HAROLD J LEAVITT
Graduate School of Business, Stanford University, Stanford, CA 94305

Professor GERALD J LIEBERMAN
Department of Operations Research, Stanford University, Stanford CA 94305

Sir FRANK MCFADZEAN
Speedbird House, Heathrow Airport (London), P.O. Box 10, Hounslow, Middlesex, TW6 2JA, U.K.

Professor RINTARO MURAMATSU
Department of Industrial Engineering, Waseda University, 4-170 Nishiokubo, Sinjuku-ku, Tokyo 160, Japan

NAOMI YAMAKI, President
Mitsubishi Space Software Co Ltd, 2-4-1, Hamamatsu-Chou, Minato-Ku, Tokyo 105, Japan.

Pergamon Press Oxford · New York · Paris · Frankfurt

AUTHOR INDEX

Aggarwal SC 161
Baker C 110
Berhold MH 275
Betts RJ 515
Bezdek RH 211
Bigham PE 183
Birchall DW 433
Booth GG 99
Briggs GH 93
Brown RV 25
Bunn DW 371, 455
Carrie AS 353
Church JG 59
Clark DJ 161
Clayton ER 191
Cooper DF 460
Corlett EN 399
Cosmetatos GP 443
Curtiss WD 378
Dar-El EM 313
Davis KR 257
De Waele M 5
Donaldson J 112
Donaldson L 109
Dukes DL 378
Dutta BK 123
Eilon S 1, 103, 201, 291, 383, 469
Elvers D 493
Engwall L 485
Felson J 237
Fisher MJ 331
Fitzgerald P 231
Flowers AD 378
Fuller JA 479
Garbin DA 331
Garrod NW 93, 374
Gentry JW 280
Gibbons M 65
Glazebrook KD 193
Glover F 145
Glover R 305
Godwin M 389
Goodwin JC Jr 493
Goodwin JS 493
Gordon IM 59
Green RH 207
Gronhaug K 501
Gross CW 531
Gummett P 65
Haley KB 369
Harding B 507
Harris TC 331
Hickson DJ 109
Hill JL 267
Hindelang TJ 267
Hindle D 370
Hollier RH 277
Hollis MS 249
Houck EC 191, 196
Hurrion RD 419
Jenkins JO 458
Johannes JD 523
Johnson LW 71
Kahalas H 53
Kaloo U 295
King WR 123
Knepley JE 331
Kynaston Reeves T 114
Lev B 282
Lockett AG 227
Mansfield R 133
McGinnis LF 325
Michael AJ 43
Mogg JM 183
Mojena R 99
Moore JM 353
Morres B 93, 374
Muhlemann AP 227
Nicol DJ 15
Noonan F 462
Nussbaum H 531
Nutt PC 33
Nuttle HLW 325
O'Brien TV 280
Parsons AT 399
Permut SE 43
Peston MH 117
Prem Vrat 277
Radford KJ 139
Rivett BHP 407
Roczniak M 353
Rodriguez JT 123
Round JI 507
Sayles L 110
Secker RJR 419
Seppänen JJ 353
Smith DG 208
Smith LD 83
Soyster AL 282
Stainton RS 284, 363
Steinberg E 97
Stewman S 341
Talmey P 305
Taylor BW III 257
Taylor JC 153
Tell B 451
Todd D 133
Towill DR 295
Tsatsos CD 91
Uhr EB 53, 173, 196
Ulvila JW 25
Watson-Gandy CDT 369
Wheeler J 133
Wiedemann P 427
Wilkes RE 173
Wyskida RM 523

TITLE INDEX

Adapting Production for Private Label Goods 363
Aggregate Planning with Inaccurate Inventory Data 97
Approach to Linear Modeling with Dispersion Surrogates 282
Bayesian Forecasting Model for House Account Sales 455
Benefits of Planning with Normative and Heuristic Approaches 33
Better than the Oracle (Editorial) 103
Capital Budgeting: A Sequential Decision Model 462
Company Structure and Market Strategy 133
Compliance Costs—The Cost of Paying Tax 389
Computers and Privacy 458
Conservatism in Bayesian Information Processing 280
Cybernetic Approach to Investment Decision Making 237
Doctor, Heal Thyself (Editorial) 383
Dynamic Programming Approach to Work Force Scheduling 378
Empirical Application of Multi-Dimensional Scaling to Advertising Pretesting 173
Empirical Examination of the Dimensions of Quality of Working Life 153
Equivalent Value Transform 275
Evaluating Time/Cost Factors of Implementation via GERT Simulation 257
Estimation of a Dirichlet Prior Distribution 371
Evaluation of Index Formulae 196
Factor Analysis—A Method for Reducing the Number of Criteria in a Multiple-Criteria Model 451
GERT Model of Discrete Probability Distributions and Sums 191

Goals and Realities in a Distribution System—The Case of Pharmaceuticals in Norway 501
Graph Theory and Computer Aided Facilities Design 353
How to Analyse Changes in Performance (Editorial) 291
Implicit Enumeration Algorithm for Solving Zero-One Goal Programming Problems 374
Influence of Aircraft Size on Airline Operating Costs 15
Input–Output Analysis and Management Accounting—An Application Concerned with Scrap Recycling 507
International Register of Theses and Dissertations 101, 197, 287, 379, 465, 541
Inventory Control by Budget 284
Large Scale Communication Networks—Design and Analysis 331
Law of the Final Cause (Editorial) 1
Management Science Application in the Courtroom: A Jury Selection Model 531
Management Science in Developing Countries: Adopting the Right Style 370
Management Science in the United States and Europe: A Decade of Change in the Literature 43
Managerial Style and the Design of Decision Aids 5
Market Share Model of Rural Telephone Service 59
Markov and Renewal Models for Total Manpower System 341
Measurement of Changes: What Criteria Do We Adopt? 399
Methodology and Teleology 91
Mixed-Model Assembly Line Sequencing Problems 313
Model of National Planning Policies 161
Modeling and Solution of Practical Management Problems through Netforms 305
Multidimension Scaling for Multiobjective Policies: A Comment (Feedback) 207
New Model of Aggregate Output Planning 267
Nuclear Medicine Patient/Instrument Scheduling Model 523
On Optimal Control and Macroeconomic Policy 117
On the Design and Control of Crisis Games 460
On the Role of Operational Research (Feedback) 208
Optimal Solutions Versus 'Good' Solutions: An Analysis of Heuristic Decision Making 479
Organisational Choice and the Woodward Thesis (Feedback) 114
Overtime Usage in a Job Shop Environment 493
Parametric Branch and Bound 145
Planning Models for Budgeting Teaching Resources 83
Planning with Multiple Objectives 427
Political Man (Editorial) 201
Postwar Strucutral and Technological Change in the American Economy 211
Pressure of Growth as an Organisational Problem 485
Productivity Drift in Extended Learning Curves 295
Project Coordinator's Problem 325
Proposal for Classification of Inventory Systems 277
Redeployment and Diversification at Harwell 65
Regression with Random Coefficients 71
Role and Needs of the Management Science Practitioner 231
Scheduling a Multi-Stage Production System with Startup Delays 183
Short-Term Foreign Exchange Risk Management: Zero Net Exposure Models 249
Social Responsibility Goal Perception 53
Some Initial Specification for a Strategic Information System 139
Some Practical Considerations on Multi-Server Queues with Multiple Poisson Arrivals 443
Some Ranking Formulae for Alternative Research Projects 193
Step-Through Simulation 25
Strategic Competitive Information Systems 123
Structural Determinism (Feedback) 109
Structural Mapping Applied to Single Value Policies 407
Student Nurse Allocation Problem: A Formulation 93
Switching Model for the Housing Decision 99
Teaching of Business Policy—A Fresh Approach 515
Travelling Salesman Distances 369
Use of Formal Inventory Control Models: A Preliminary Survey 227
Visual Interactive Simulation—An Aid to Decision Making 419
Vroom-Yetton Leadership Model (Editorial) 469
Work System Design and the Quality of Working Life 433

ol. 6
1978

Aims and Scope

The purpose of OMEGA is to provide a vehicle for reporting developments in management science, operational research and managerial economics, including research results and applications. Published material ranges from original contributions to review articles describing the state of the art in specific areas, together with shorter critical assessments of particular management techniques. Other features of the journal are the Memoranda section for short communications, Feedback (a correspondence column) and a register for theses and dissertations.

The journal is intended to serve the needs of practising managers, specialists in management services, operational research workers and management scientists, management consultants, academics, students and research personnel.

Contributions may be made in any of the following categories:

1. *State of the art.* These papers would normally be more than 5000 words long, devoted to an interpretation and evaluation of existing knowledge in a particular field or function of management. Such papers should aim to be valuable both to practising managers and also to research workers and should therefore include a detailed bibliography.

2. *Theory.* This category caters for papers in specialised fields, 1500–5000 words long, which report research results and new theoretical developments.

3. *Applications.* These papers describe interesting applications of management science techniques and should typically be 1500–5000 words in length.

4. *Memoranda.* This section is devoted to short contributions, preferably under 600 words, to preliminary communications for new ideas and propositions, technical notes, brief descriptions of internal research reports or theses/dissertations, extension or critique of widely accepted concepts, brief accounts of applications and implementation. The intention is to publish such memoranda with as little delay as possible. Items for Memoranda Section should have the minimum number of diagrams (preferably none if possible) and lists of references should also be pared to he absolute minimum.

5. *Feedback.* Correspondence is invited on issues of significance in the field of management. This need not necessarily relate to material previously published in OMEGA.

6. *International Register of Theses and Dissertations.* Information for inclusion in this Section should be presented in the style of the Journal.

Annual Subscription Rates (1979)

Published bi-monthly. For libraries, university departments, government laboratories, industrial and other multiple-reader institutions U.S. $115.50 per annum (including postage and insurance). *Specially reduced rates to individuals*: In the interests of maximising the dissemination of material published in this important international journal we have established a two-tier price structure. Any individual whose institution takes out a library subscription may purchase a second or additional subscription for their use at the much reduced rate of U.S. $30.00 (£12.50) *per annum.* All subscription enquiries should be addressed to: The Subscription Fulfilment Manager, Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England.

Microform Subscriptions and Back Issues

Back issues of all previously published volumes are available in the regular editions and on microfilm and microfiche. Current subscriptions are available on microfiche simultaneously with the paper edition and on microfilm on completion of the annual index at the end of the subscription year.

Publishing Office

Pergamon Press Ltd., Hennock Road, Marsh Barton, Exeter, Devon EX2 8RP, England.

Subscription and Advertising Offices

Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England.
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.

Copyright 1978 Pergamon Press Ltd.

It is a condition of publication that manuscripts submitted to this journal have not been published and will not be simultaneously submitted or published elsewhere. By submitting a manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microform or any other reproductions of similar nature and translations. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the copyright holder.

U.S. Copyright Law applicable to users in the U.S.A.

The Article Fee Code on the first page of an article in this journal indicates the copyright owner's consent that in the U.S.A. copies may be made for personal or internal use provided the stated fee for copying beyond that permitted by Section 107 or 108 of the United States Copyright Law is paid. The appropriate remittance should be forwarded with a copy of the first page of the article to the Copyright Clearance Centre Inc. PO Box 765, Schenectady, NY 12301. If a code does not appear copies of the article may be made without charge, provided permission is obtained from the publisher. The copyright owner's consent does not extend to copying for general distribution, for promotion, for creating new works or for resale. Specific written permission must be obtained from the publisher for such copying. In case of doubt please contact your nearest Pergamon office.

Printed in Great Britain by A. Wheaton & Co. Ltd., Exeter

Vol. 6
1978

OMEGA

The International Journal of Management Science

Volume 6, Number 1, 1978

CONTENTS

SAMUEL EILON	1	Editorial—The Law of the Final Cause
MARTIN DE WAELE	5	Managerial Style and the Design of Decision Aids
DAVID J NICOL	15	The Influence of Aircraft Size on Airline Operating Costs
JACOB W ULVILA, REX V BROWN	25	Step-Through Simulation
PAUL C NUTT	33	The Benefits of Planning with Normative and Heuristic Approaches
AJ MICHEL, SE PERMUT	43	Management Science in the United States and Europe: A Decade of Change in the Literature
HARVEY KAHALAS, ERNEST B UHR	53	Social Responsibility Goal Perception
JG CHURCH, IM GORDON	59	Market Share Model of Rural Telephone Service
PHILIP GUMMETT, MICHAEL GIBBONS	65	Redeployment and Diversification at Harwell
LW JOHNSON	71	Regression with Random Coefficients
LD SMITH	83	Planning Models for Budgeting Teaching Resources
	89	**Memoranda**
HIS EXCELLENCY CONSTANTINE D TSATSOS	91	Methodology and Teleology
B MOORES, NW GARROD, GH BRIGGS	93	The Student Nurse Allocation Problem: A Formulation
EARLE STEINBERG	97	Aggregate Planning with Inaccurate Inventory Data
GG BOOTH, R MOJENA	99	Switching Model for the Housing Decision
	101	International Register of Theses and Dissertations

Special Notice

To avoid simultaneous, multiple submissions to journals having similar interests, the Editors of *Management Science*, the *Operational Research Quarterly*, *Operations Research*, *Omega* and *Naval Research Logistics Quarterly* have developed a system of exchange submission lists.

Because refereeing capability is a scarce resource, we cannot tolerate the duplication of effort that follows from multiple submissions. We hope that the elimination of multiple submissions will make the refereeing process more efficient, and thereby decrease the time required for refereeing. This is a major objective of all five publications. Other journals with similar interests may soon be included in this information interchange.

Vol. 6
1978

ol. 6
978

OMEGA

The International Journal of Management Science
Volume 6, Number 2, 1978

Contents

SAMUEL EILON	103	Editorial—Better than the Oracle?
	109	Feedback
MAURICE H PESTON	117	On Optimal Control and Macroeconomic Policy
WILLIAM R KING, BIPLAB K DUTTA, JAIME T RODRIGUEZ	123	Strategic Competitive Information Systems
R MANSFIELD, D TODD, J WHEELER	133	Company Structure and Market Strategy
KJ RADFORD	139	Some Initial Specifications for a Strategic Information System
FRED GLOVER	145	Parametric Branch and Bound
JAMES C TAYLOR	153	An Empirical Examination of the Dimensions of Quality of Working Life
SC AGGARWAL, DJ CLARK	161	A Model for National Planning Policies
RE WILKES, EB UHR	173	An Empirical Application of Multidimensional Scaling to Advertising Pretesting
JM MOGG, PE BIGHAM	183	Scheduling a Multi-Stage Production System with Startup Delays
	189	**Memoranda**
ER CLAYTON, EC HOUCK	191	GERT Model of Discrete Probability Distributions and Sums
KD GLAZEBROOK	193	Some Ranking Formulae for Alternative Research Projects
EB UHR, EC HOUCK	196	Evaluation of Index Formulae
	197	International Register of Theses and Dissertations

omega

The International Journal of Management Science

Volume 6, Number 3, 1978

Contents

SAMUEL EILON	201	Editorial—Political Man
	207	Feedback
ROGER H BEZDEK	211	Postwar Structural and Technological Change in the American Economy
AP MUHLEMANN, AG LOCKETT	227	The Use of Formal Inventory Control Models: A Preliminary Survey
PADDY FITZGERALD	231	The Role and Needs of the Management Science Practitioner
J FELSEN	237	Cybernetic Approach to Investment Decision Making
MARTHA S HOLLIS	249	Short-Term Foreign Exchange Risk Management: Zero Net Exposure Models
BERNARD W TAYLOR III, K ROSCOE DAVIS	257	Evaluating Time/Cost Factors of Implementation via GERT Simulation
THOMAS J HINDELANG, JOHN L HILL	267	A New Model for Aggregate Output Planning
	273	**Memoranda**
MH BERHOLD	275	An Equivalent Value Transform
RH HOLLIER, PREM VRAT	277	A Proposal for Classification of Inventory Systems
TV O'BRIEN, JW GENTRY	280	Conservatism in Bayesian Information Processing
AL SOYSTER, B LEV	282	An Approach to Linear Modeling with Dispersion Surrogates
RS STAINTON	284	Inventory Control by Budget
	287	International Register of Theses and Dissertations
	I	Announcement

Vol. 6
1978

ol. 6
978

omega

The International Journal of Management Science
Volume 6, Number 4, 1978

Contents

SAMUEL EILON	291	Editorial—How to Analyse Changes in Performance
DR TOWILL, U KALOO	295	Productivity Drift in Extended Learning Curves
RANDY GLOVER, PAUL TALMEY	305	Modeling and Solution of Practical Management Problems through Netforms
EZEY M DAR-EL	313	Mixed-Model Assembly Line Sequencing Problems
LF MCGINNIS, HLW NUTTLE	325	The Project Coordinator's Problem
MJ FISCHER, DA GARBIN, TC HARRIS, JE KNEPLEY	331	Large Scale Communication Networks —Design and Analysis
SHELBY STEWMAN	341	Markov and Renewal Models for Total Manpower System
ALLAN S CARRIE, JAMES M MOORE, MAREK ROCZNIAK, JOUKO J SEPPÄNEN	353	Graph Theory and Computer Aided Facilities Design
RS STAINTON	363	Adapting Production for Private Label Goods
	367	**Memoranda**
KB HALEY, CDT WATSON-GANDY	369	Travelling Salesman Distances
DON HINDLE	370	Management Science in Developing Countries: Adopting the Right Style
DEREK W BUNN	371	Estimation of a Dirichlet Prior Distribution
NW GARROD, B MOORES	374	An Implicit Enumeration Algorithm for Solving Zero-One Goal Programming Problems
A DALE FLOWERS, DAVID L DUKES, WALTER D CURTISS	378	A Dynamic Programming Approach to Work Force Scheduling
	379	International Register of Theses and Dissertations

omega

The International Journal of Management Science

Volume 6, Number 5, 1978

Contents

SAMUEL EILON	383	Editorial—Doctor, Heal Thyself
MICHAEL GODWIN	389	Compliance Costs—The Cost of Paying Tax
EN CORLETT, AT PARSONS	399	Measurement of Changes: What Criteria Do We Adopt?
PATRICK RIVETT	407	Structural Mapping Applied to Single Value Policies
RD HURRION, RJR SECKER	419	Visual Interactive Simulation—An Aid to Decision Making
PAUL WIEDEMANN	427	Planning with Multiple Objectives
DAVID W BIRCHALL	433	Work System Design and the Quality of Working Life
GEORGE P COSMETATOS	443	Some Practical Considerations on Multi-Server Queues with Multiple Poisson Arrivals
	449	**Memoranda**
BERTILL TELL	451	Factor Analysis—A Method for Reducing the Number of Criteria in a Multiple-Criteria Model
DEREK W BUNN	455	A Bayesian Forecasting Model for House Account Sales
JO JENKINS	458	Computers and Privacy
DALE F COOPER	460	On the Design and Control of Crisis Games
FRANK NOONAN	462	Capital Budgeting: a Sequential Decision Model
	465	International Register of Theses and Dissertations

Vol. 6
1978

ol. 6
978

omega

The International Journal of Management Science
Volume 6, Number 6, 1978

Contents

SAMUEL EILON	469	Editorial — The Vroom–Yetton Leadership Model
JACK A FULLER	479	Optimal Solutions Versus 'Good' Solutions: An Analysis of Heuristic Decision Making
LARS ENGWALL	485	The Pressure for Growth as an Organizational Problem
JAMES C GOODWIN JR, DOUGLAS ELVERS, JACK S GOODWIN	493	Overtime Usage in a Job Shop Environment
KJELL GRØNHAUG	501	Goals and Realities in a Distribution System: The Case of Pharmaceuticals in Norway
B HARDING, JI ROUND	507	Input-Output Analysis and Management Accounting: An Application Concerned with Scrap Recycling
RJ BETTS	515	The Teaching of Business Policy—A Fresh Approach
JAMES D JOHANNES, RICHARD M WYSKIDA	523	A Nuclear Medicine Patient/Instrument Scheduling Model
CHARLES W GROSS, HARVEY NUSSBAUM	531	A Management Science Application in the Courtroom: A Jury Selection Model
	541	International Register of Theses and Dissertations
	i	Announcement

OME. 6/6—A

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 1–3, 1978. Pergamon Press. Printed in Great Britain

EDITORIAL

The Law of the Final Cause

IN A NOTE published in this issue (p. 91) His Excellency Constantine Tsatsos writes, not in his capacity as the President of the Hellenic Republic, but as a former professor of social philosophy, and he attempts to draw attention to the relationship between methodology and goal seeking in any human activity, particularly in the field of management. Since methodology is a plan, or a series of procedures, for marshalling thoughts and actions, it must have a purpose: "method is a course consciously followed", as Tsatsos puts it, and it is followed in the pursuit of "the final cause", the cause that ultimately motivates the individuals concerned to persist in their search for it. The final cause and the final goal thus become one and the same, since the goal is the reason for the search to be undertaken in the first place, and the methodology is simply the means envisaged for achieving it.

Tsatsos uses the analogy of a captain of a ship at sea; he collects information from various sources and employs equipment at his disposal for the sole purpose of steering his ship to its destination. All his actions must, in the final analysis, be interpreted and judged against that goal. Similarly, in other human and managerial affairs, all activities need to be organized with the final cause in mind, so that when in a given situation the goal is clear to all concerned, the only issue that may be in dispute is the means of attaining it.

It is in this area, namely in the development of effective methodology, that management science appears to have made its greatest contribution and where immense potential for further work remains "to rationalize action in order to manage it". The purpose of model building is thus seen as a way of guiding the captains of industrial and business systems to navigate as smoothly and effortlessly as possible towards their goals. A complication that Tsatsos draws attention to is that "life calls us to action even without complete knowledge", so that decisions have to be made with limited, irrelevant, or even biased information, within strict time constraints. Models which purport to describe reality under conditions of perfect information may not be much use to decision makers under other conditions, where disputes about the efficacy of alternative methods may not be easy to resolve. Even so, the task of management science in evaluating the performance of various methods, in perfecting new ones, and in trying to draw general conclusions to aid the manager, is generally acknowledged as an important function to be performed for many years to come. The day when all questions about methodology will be settled and when the manager can 'plug in' the most appropriate and efficient procedure to handle a given problem is certainly not yet in sight.

What intrigues me about this scenario is not so much the notion that the development of methodology is thought to be independent of the process of determining what the goal is, but the implication that management scientists are supposed to take that goal for granted. In other words, goal setting

(policy) and methodology (strategy) are envisaged as separate roles performed in tandem, based on the proposition that the goal and the means for achieving it are not necessarily linked to each other through a mechanism of recurring feedback, so that in certain situations it may be possible to engage in a goal setting exercise in the absence of any knowledge of the information and the decision processes that are, or can be, employed by those charged with the task of achieving the goal. It is a scenario in which the management scientists are denied the opportunity—I would go as far as to say the responsibility—of questioning and contributing to the delineation of goals. I have commented elsewhere on the fact that the role of the OR (operational research) analyst is often perceived as that of a technician, and I pointed out the implications that such a confined role can have on the way an OR group is placed in an organizational structure.[1] The comments were made in the context of OR being employed on tactical versus strategic problems in a business enterprise, but similar observations may be advanced about problems solving versus goal setting.

What is the relevance, then, of the analogy of the ship's captain at sea? It attempts to demonstrate that separation of the goal from methodology can sometimes be made. The captain is told by the owner of the ship to go from *A* to *B*; it is not his function to ask why *B* is chosen as a destination, whether the voyage will be profitable, or whether—out of a number of possible voyages—it is the best to undertake. It is not his function to ask these questions, because he is a technician whose expertise lies in steering ships, not in conducting financial analyses. He may not even care two hoots as to which destination he is ordered to go to (leaving aside the different attractions that he may find in each port), and his activities in steering the ship (his methodology) remain essentially the same whatever the destination.

For management scientists who see themselves as technicians this is a perfectly valid analogy, in that it provides them with a comfortable line of demarcation between the responsibility of deciding where a system should be going and determining how it should get there. But problems begin to emerge when the 'final cause' is not obvious, when several causes can be clearly identified, and when reconciliation between causes is not easily demonstrable. In short, a multi-goal system poses not only questions as to how the goals should be defined and ordered, but about the kind of methodology that should be developed and used in steering the system for the purpose.

Another point that needs to be made is that the term 'final' in the concept of the final cause raises inevitable difficulties. Immediate goals can always be interpreted as milestones on the road to other goals, and those in turn lead to further goals, more remote and more fuzzy to define or to measure. A series of questions soon reveals to the decision maker how difficult, or even impossible, it is for him to identify a (or the) true final cause:

— Why do you make product *X*? What is the purpose of its production?
— The purpose is to sell it in the market.
— Why do you want to sell it?
— Because it meets a need; besides, I want to cover my costs and make a profit.
— Why do you want to make a profit?
— To pay shareholders, to expand, to improve the prospects of my employees.
— Why?

[1] See "Technician or adviser?", Editorial, *Omega* **5** (1), 1–6 (1977).

Vol. 6
1978

— To beat the competition, to survive, to attain security, well-being and happiness.
— Why? Why?

So, what is the final cause? Is it survival, fulfilment, power, happiness? The further we go along this road, the more we realise that motivation is a complex theme and that the final cause cannot merely be described in terms of a hierarchy of needs. The immediate milestones are relatively easy to describe, but those further away do not lend themselves to quantification and measurement, nor is it possible to relate them to a time scale. For example, a goal associated with production or profit can be specified in the form of certain targets of volume and money, say by a given date. It should then be possible (subject to given conventions of how variables in the system are to be measured) to verify whether on the day in question the goals have been achieved or not. But more remote and amorphous goals, such as fulfilment or happiness, are at best only loosely time-dependent, and even if reliable measures were readily available to indicate whether a state of happiness has improved from year 1 to year 2, even if we could resolve the problem that an important state of happiness for an organization (however it is defined) may be detrimental to the state of happiness of an individual, it would be impossible to relate such changes to particular and immediate tasks that organizations or individuals have to perform.

It is for these reasons that the law of the final cause must be regarded as too simplistic a proposition for explaining the activities undertaken by enterprises, the organization structure adopted, the allocation of duties and responsibilities to groups and individuals, the procedures developed for planning and accountability. To interpret all these intricacies (which constitute what may be regarded as the 'methodology' of the enterprise) as the manifestations of an overwhelming presence of a distant and somewhat obscure final cause would be entenable; it may be argued that only in relation to more immediate goals can activities be planned and executed, and that only in terms of goals that can be quantified is it possible to make judgements about the adequacy or otherwise of organizational performance.

And so, if measures of performances are to be set in conjunction with targets or yardsticks, and if one single measure is not universally acclaimed for the purpose, the answer to the problem of what goals to pursue is often far from self-evident. Method and analysis are needed as much for setting goals as they are for determining means for attaining them, and this is not a responsibility that management scientists can or should be allowed to shirk.

SAMUEL EILON
Chief Editor

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 5–13, 1978. Pergamon Press. Printed in Great Britain

Managerial Style and the Design of Decision Aids

MARTIN DE WAELE

University of California, Berkeley

(*Received May 1977; in revised form June 1977*)

A science of managerial decision making and of aids to management describes the nature and dynamics of decision making. This article proposes one possible descriptive framework of decision making styles for marketing and product managers. It has consequences for both communication between designer and manager and the design of the aids. A decision aid is, per definition, an aid designed to assist a manager of a particular style in making decisions. Different kinds of decision aids can be conceived to match different styles, along certain criteria. Several linkage criteria between aids and styles will be briefly explored here.

SCIENCE, OBSERVATION AND IMPROVEMENT

IT HAS become apparent that when a scientist begins to look into the essence of management and decision making, whether it is the management of a product, a company, a nation, or indeed any project, he soon realizes that mathematical formulation of problems in order to derive better solutions often implies an extremely one-sided view of the whole situation. The ensuing astonishment on the part of the scientist, frequently characterized by the notion of implementation problem, may be due to the fact that he did not observe what happens in actual decision making situations. He did not observe first what the whole situation is like, that is, what managers—for whom he is developing an aid—actually do in the problem situation they encounter. One could say he has put his cart before the horse in the enthusiasm of his intellectual powers. In many cases he may not even have observed his own decisions.

Observation involves description and categorization. There is no reason why the social sciences in general should omit the preliminary phenomenological schemes based mainly on pure observation, which seem to have been so useful for the arrangement of facts in disciplines like chemistry or particle physics [9]. Only after such work can one give due attention to dynamic properties and, finally, to the manipulation of this observed material for whatever purpose we choose, for example, making better decisions. This cry for description is even more acute for human behavior than for chemistry or physics, because we all tend to think about behavior on the basis of our own psychological make-up. Frequently the 'oughts' that we utter are based on the biases of our own individuality. This holds not only for statements like "if you want A, then logically you ought to do B", but also for "the best way to solve A is through steps XYZ". The way a particular problem can be solved is not unique, and accordingly the 'best' way may depend on our own individuality. Only after observing the situations and managerial behavior can the scientist begin to think about how the process of management and decision making can be enhanced. Naturally this is a venture for both manager and scientist.

The logical result of the fact that there are different individualities is that the adequacy of any decision aid will differ accordingly. It is therefore reasonable to expect the decision maker to evaluate whether something constitutes an aid for him or not because, in his decision making capacity, not only does he experience what he is doing, but it is he, not the scientist, who carries the final responsibility for what he does. However, should acceptance of the aid by the manager be the only criterion for its adequacy? To say that it should implies a moral choice: it puts the entire responsibility in the manager's hands. In fact, however, does

this take away all responsibility from the scientist? In designing an aid he cannot avoid struggling with criteria for improvement: enhancing efficiency, reducing fatigue, reducing decision anxiety, enhancing communication between manager and vice-president, or indeed anything. What, for example, is the role of the scientist if a manager is fired because he tried to use models and made bad decisions? This suggests that a greater effort will be required of the scientist to transcend his own individuality to develop aids responsive to the manager's needs.

In his study on the nature of managerial work Mintzberg summarized previous descriptive research on management and characterized the work of chief executives in terms of roles [20]. In this paper the focus is on the informational and decision making roles of marketing and product managers, in the context of relatively unstructured problems. I further refer to both these roles as decision making. The unstructured nature of problems may mean many things. For example, the number of alternatives is open-ended, relationships between variables are complex and not well-known, the formulation of the problem is extremely difficult, some symptom is a problen in one case but not in another, uncertainty prevails in the face of competitive reactions and initiatives, consumer reactions, and legal reactions, immediacy of remedial action is required for certain problems, the execution part of any decision is of utmost importance rendering decision and execution inseparable.

In the remainder of the paper I describe what decision making styles may mean, some structural properties of the functions which are components of one possible style framework, the style framework itself, its consequences for communication between manager and designer (management scientist), and its consequences for design criteria for decision aids.

STYLES OF DECISION MAKING: EXPLORATION

Two basic premises underly the description that will follow in the next sections. The first one is that people are essentially different from each other in the way they go about decision making. The second one is that they actually tend to focus on different kinds of problems, in one and the same job. Several authors have found evidence for one or both of these statements.

Regarding the first issue, a number of examples are available from previous research. In the management area, McKenney and Keen found important differences in the way individuals of different styles approach problems and data. In particular, they found that 'systematic' thinkers tend to plan methodically for solving a problem and conduct an ordered search for additional information, while 'intuitive' thinkers tend to jump from one step in analysis or search to another and back again and consider a number of alternatives and options simultaneously [13]. Stabell identified certain behavioral differences in the way managers went about making investment decisions on the basis of their scores on Kolb's learning style inventory. Within one department of a bank, for example, he found that managers of a Concrete Experience style cited people as more important information sources, and sought services that would give them specific recommendations which they could accept or reject, while managers of an Abstract Conceptualization style tended to prefer printed material as more important information sources, and sought information which they could analyze themselves in order to choose an investment. In addition, they tended to use more information sources than Concrete Experience style managers [22]. In the medical area, Leaper and his collegues observed a number of diagnoses by clinicians and found a number of differences among individuals leading them to conclude that 'the diagnostic process' does not exist [16].

Regarding the second issue, different styles focusing on different problems within the same job, McKenney and Keen again relate their findings that the 'systematic' individuals tended to prefer program-type problems, while the 'intuitive' individuals tended to prefer open-ended ones, especially those that required ingenuity or opinion [17]. An interesting corollary is that similar differences were observed in career preferences [3]. A fairly conspicuous illustration of this we find in presidential decision making in the United States: even though the job remains essentially the same, each president

Vol. 6
1978

tends to mold the job and the problems he deals with into his own frame of mind, and has his own special preferred ways of dealing with them. Compare, for example, the general decision making images that the last five presidents have left with us. And in our own lives, have we not all experienced difficulties understanding the way a colleague thinks? Or have we not been surprised, on occasions, to see that our way of thinking differs radically from that of somebody else, even though we may arrive at essentially the same conclusion?

A good starting point to explore such general characteristics of managerial decision making is Jung's theory of psychological types.[1] No one can do justice in a brief paper to the enormous richness of his framework. The reasons for starting with such a general typology are manifold. First, so little descriptive work has been done so far in managerial decision making that I think that at this stage we can learn more by going from the general to the particular, rather than vice versa. It is sometimes difficult to expand upon an originally narrow focus. One feels this, for example, with the studies of cognitive complexity in implementation research, despite the substantial contributions which these studies have made. Second, whereas other typologies emphasizing personality characteristics have tended to have either an aura of good–bad notions and determinism about them (for example, [15]), or an aura of less mature–more mature notions (for example [14]), Jung emphasizes that no style is better than any other and to me, as to Kilmann and Mitroff [13], this is an extremely important idea. Third, there is an easy, although not perfect, way to find out about an individual's preferred functions by means of a questionnaire. One such questionnaire is the Myers-Briggs Type Indicator which has a large amount of research back-up and is fairly easy to administer [3]. Finally, the theory has a firm ground in experience and observation.

[1] While the word 'type' tends to have a static connotation, it is important to note that Jung himself used it as a more fluid and dynamic concept. Therefore I have chosen to use the term 'style' where he used 'type'.

SOME STRUCTURAL CHARACTERISTICS OF THE FUNCTIONS OF THE MIND

According to Jung's theory, people have four main functions at their disposal: they gather data with their perceptive functions, sensation or its opposite, intuition. These he calls the irrational functions. The data are then processed by the rational, assessing or judging functions: thinking, at one extreme, and feeling at the other extreme. The way the four functions are conceived is fully described elsewhere [11, 12, 19], so I will only repeat some of the essentials here and bring some of the sources together.

"Thinking is the function which seeks to apprehend the world and adjust to it by way of thought or cognition (logical inferences)" [11]. The function of feeling, on the other hand, apprehends the world through evaluations along the lines of liking or disliking, good or bad, pleasant or unpleasant, or value judgments. Both these functions are called rational, because both work with evaluations and judgments: thinking evaluates through cognition from the standpoint of true–false, feeling through judgments from the standpoint of liking–disliking. Feeling is a subjective process. It is a kind of judgment differing from intellectual judgment in that its aim is not to establish conceptual relations but to set up a subjective criterion of acceptance or rejection. "As determinants of behavior, these two basic functions are mutually exclusive at any given time" [11]. Either the one or the other predominates, because they are both evaluative.

The other two functions, sensation and intuition, are called irrational functions "because they circumvent the ratio and operate not with judgments but with mere perceptions which are not evaluated or interpreted. Sensation perceives things as they are and not otherwise. Intuition also 'perceives' but less through the conscious apparatus of the senses than through its capacity for an unconscious 'inner' perception of the inherent potentiality of things" [11]. The sensation function, for example, will note all the details of a historical event, but tend to disregard the general context in which it is set; the intuitive function, on the other hand, will pay little attention to the details but will have no difficulty in discerning the inner mean-

ing of the event, its possible implications and effects. Thus, these two functions "are just as antithetical and mutually exclusive as the first pair, thinking and feeling; they cannot operate simultaneously" [11].

All men and women possess all four functions. But experience has shown that usually one function predominates [3, 12, 23]. Somehow our constitution, education and environment has helped us develop one function in a more pronounced way. This one function "is always available to the individual's conscious will" [21]. When using that function predominantly becomes a habit, we speak of a style. That function is also called the individual's superior function. Its opposite becomes the inferior function; it is typically less differentiated, less well developed and is subordinated to the superior function. One of the two other functions is usually available as an auxiliary function. That is possible because it is a function of a different nature than the superior function, although it is not antagonistic to it. It serves the superior function. Usually the superior and auxiliary functions work well together [21].

In addition to those basic functions, Jung distinguishes between two ways of reacting to outer and inner experience, extraversion and introversion. Extraversion is an 'outward-turning' of psychic energy, and implies a 'positive movement of subjective interest towards the object'. Similarly, introversion is an 'inward-turning' of psychic energy. "Interest does not move towards the object but withdraws from it into the subject...the subject is the prime motivating factor and the object is of secondary importance" [12]. When one of those ways becomes habitual, again, we can talk about a style. The preference between extraversion and introversion affects the individual's choice as to whether to direct perception and judgment mainly upon the environment or the world of ideas [3].

It should be emphasized that pure styles do not exist. Furthermore, this framework is merely intended as a helpful working hypothesis for our purpose which, eventually, is to examine design features of decision aids. In that sense it will suffice to talk about the structural characteristics of the functions and to omit most of their dynamic characteristics, which are related to the psychological growth of the individual. The next section is devoted to what these functions may mean in the context of decision making styles of marketing and product managers.

PREFERRED FUNCTIONS AND DECISION MAKING STYLE

Without any doubt all four functions have something to offer in the area of marketing management. Indeed, this area not only involves pure decision making somewhere in the minds of managers, but it involves information gathering (from reports and from people), monitoring the market, executing decisions, and evoking enthusiasm both on the part of higher management and on the part of the executors. It also involves thinking about new possibilities and about possible consequences of actions, and certainly it involved all kinds of value, ethical, political, and power considerations on the part of the manager. In other words, management is not a machine-like activity for the manager, but one in which his whole personality is involved with his individual experiences as well as the operating procedures and atmosphere of his organization. Although certain types of careers undoubtedly tend to attract certain styles of individuals [3], in view of the pervasiveness of management functioning and, in particular, the wide variety of activities involved in marketing management, there is no *a priori* reason why we should not find all kinds of styles among marketing and product managers.

In the remainder of this section a number of styles which we may expect among managers will be described, along with some of their characteristics. It should be emphasized also, at the outset, that contrary to what Hellriegel and Slocum seem to imply [10], the purpose of classifying decision making styles in this fashion is not to force people into slots and to argue that such or such a style is more suited for this or that kind of a job. Looking at a particular kind of job, that of marketing or product management, the question is rather: given that we find different styles there, how are some of the decision making characteristics of these managers different, and what are the implications for the design of decision aids.

For reasons of clarity and in keeping closely to the structural ideas of the theory, let us assume that each individual tends to use habi-

tually one superior function and one auxiliary function. These are the functions which he is reasonably comfortable with. He has most difficulty with his inferior function, which is the opposite of his superior function. We then have several possible styles, each of which can be extraverted or introverted:

Sensation–Thinking, with inferior Intuition
Thinking–Sensation, with inferior Feeling
Intuition–Thinking, with inferior Sensation
Thinking–Intuition, with inferior Feeling
Sensation–Feeling, with inferior Intuition
Feeling–Sensation, with inferior Thinking
Intuition–Feeling, with inferior Sensation
Feeling–Intuition, with inferior Thinking

When a perceiving function (sensation or intuition) is superior, we can expect particular characteristics in the way a manager approaches his decision making. Whatever such a style does or does not do "is based not on rational judgment but on the sheer intensity of perception" [12]. His perception is directed to events as they happen, no selection being made by judgment. The advantage over a judging style (thinking or feeling) is clear when one considers that some objective events conform to law and some are accidental. "In so far as they conform to law, they are amenable to reason; in so far as they are accidental, they are not. But merely because they subordinate judgment to perception, it would be quite wrong to regard them as 'unreasonable'. It would be truer to say that they are in the highest degree empirical. They base themselves exclusively on experience—so exclusively that, as a rule, their judgment cannot keep pace with their experience" [12].

When sensation is the superior function, those objects or subjective elements that excite the strongest sensations will be decisive for the individual's psychology and actions, whether they are compatible with rational judgment or not [12]. The style is oriented amid the flux of events not by rational judgment but simply by what happens. "He does not always stop to take a 'thought-out' position on what he perceives nor does he react strongly with his feelings. But he can describe a room he has been in with accurate detail and has no trouble finding his way when he has once traveled over the terrain" [21]. "The judging, rational functions are subordinated to the concrete facts of sensation. The function most repressed is naturally its opposite, intuition, the function of unconscious perception" [12]. Feeling as an auxiliary function will add the capacity for sympathetic awareness of people and enthusiasm to this style, while thinking as an auxiliary function will tend to add impersonal analytical ability [3].

However, when intuition is the superior function, although the same marked dependence on perceived situations—objective or subjective—is manifest, it is different from the dependence of the sensation style. "The intuitive is never to be found in the world of accepted reality-values, but has a keen nose for anything new and in the making. Because he is always seeking out new possibilities, stable conditions suffocate him" [12]. Thinking, feeling and sensation are largely repressed, but sensation is the one most affected because, as the conscious function, it "disturbs the intuitive's clear, naïve perception with sensory stimuli, which direct attention onto the superficial, beyond which intuition tries to peer" [21]. Again, feeling as an auxiliary tends to add a capacity for sympathetic awareness of people, while thinking will tend to add capacity for analysis [3].

When a judging, rational function (thinking or feeling) is superior, then again we may expect certain peculiar ways of behaving which form the style of those managers. "Their life is, to a great extent, subordinated to rational judgment" [12]. The rationality that characterizes the conscious conduct of life in both these styles "involves a deliberate exclusion of everything irrational and accidental. Rational judgment, in such a psychology, is a force that coerces the untidiness and fortuitousness of life into a definite pattern, or at least tries to do so" [12]. A definite choice is made from among all the possibilities it offers, only the rational ones being accepted; but on the other hand, the perceptive functions are restricted. This restriction is not absolute: the functions of sensation and intuition exist but their products are subject to the choice made by rational judgment. "It is not the intensity of a sensation as such that decides action, for example, but judgment" [12]. Rational judgment may be based on objective or on subjective data.

When thinking is the superior function, the life of the individual is "mainly governed by reflective thinking, so that every important

action proceeds or is intended to proceed from intellectually considered motives" [12]. The person of this style will endeavour to make all his activities dependent on intellectual conclusions, which are oriented by objective data or by his subjective foundation. He elevates his idea, "or an objectively oriented formula into the ruling principle not only for himself but for his whole environment. Everything that agrees with this formula is right, everything that contradicts it is wrong, and anything that passes by it indifferently is merely incidental" [12]. His formula is "made into a universal law which must be put into effect everywhere all the time, both individually and collectively" [12]. The feeling function is most affected by conscious inhibition since it is opposite to the rigid intellectual formula. "The end justifies the means" and dogmatism of the intellectual formula are attitudes which can be brought about by the inferior feeling function. The other two functions may be more or less available to consciousness. When sensation comes as an auxiliary function it will tend to add realism, observation, respect for detail and reliance on experience. Intuition will tend to add insight and ingenuity, and a grasp of the complex [3].

When feeling is the superior function, every conclusion, however logical, that might lead to a disturbance of feeling is rejected at the outset. In other words, all thinking is completely subordinate to feeling values. It is not surprising that this style dislikes thinking. Such a style has tremendous capacity for objectively feeling another person's situation. Wherever something does not function properly he realizes it at once. He sees the importance or the value of what should be done [23]. Again, sensation as an auxiliary function will tend to add reliance on experience, realism and observation, while intuition will tend to add ingenuity, insight and a grasp of the complex [3].

It is fairly obvious that, even in the same job of marketing/product management, the style of different persons can vary substantially, particularly in the way data are assimilated and the way decisions are made habitually. In fact, the individual himself and his decision making style are inseparable. This should have implications for the relation between the manager and the management scientist. On the personal level, the issue is one of communication; on the technical level the issue is one of design of the decision aid. In practice the two are certainly very much related but for the sake of clarity we have to consider them separately in the next sections.

COMMUNICATION BETWEEN MANAGER AND MANAGEMENT SCIENTIST

In their classic article, Churchman and Schainblatt argued for mutual understanding between manager and management scientist [5], but some people have pointed to the lack of operationalization of that concept. The foregoing notions provide a way to operationalize one possible meaning of the concept.

Let us first take into consideration that the differences we have alluded to in the preceding sections are real; second, that there is no reason to assume that one style is better than any other; and third, that for our purposes it is not necessary to know how these differences come about in any particular case: they exist and they are a fundamental characteristic of being. There are basically, then, two ways of reacting to the problem of differences. One is to ignore it and to go away from it, feeling that different styles cannot understand each other. That is the pessimistic, negative way. The other is to recognize the complementarity and richness of the differences, feeling that each style can learn from the other style, and make his own experience of the world more complete. That is the optimistic, positive way. The first one carries the image of decay, the second one carries the image of growth.

Earlier we have made the proposition that marketing managers generally have different styles from each other, and the same idea might be applied to management scientists as well. The foregoing theory would lead to the proposition that style differences between managers and management scientists have implications for the communications between them. It is plausible to say that communication will be enhanced to the extent that each style realizes what style among possible other styles he himself has come to prefer. The process consists of becoming aware of the phenomenon of styles and gradually coming to respect the differences. At the same time their richness is revealed. The process is certainly not easy, because the tendency to feel that one's style

is the best is very strong and present in everyone. Yet it is intuitively clear that effective communication requires that kind of mutual understanding and respect. "Ordinarily when we as individuals find ourselves in uncongenial discussions, we are forced to come to terms with people whose typology is quite different from our own. They will see things one way and be quite convinced that they are correct, while we will see things in another way and be just as convinced. There can never be any reconciliation between opposing views unless there is first a recognition that matters may be seen in other terms than that of correctness or error. To realize that they are simply seen from varying viewpoints may be the beginning of understanding. For always there is one set of facts which is regarded by different people in different ways" [21].

Even if a manager and a management scientist are persons of the same style, however, that does not necessarily mean they will get along very well. But then the reason for the disagreements will probably be a different focus of interest and different specific knowledge. An extreme example of that occurs when two university professors, say of the same dominant thinking style, have opposing theories, each claiming his own theory is the correct one and the other is wrong, and sticking to his arguments. Even the same styles require some mutual understanding.

DECISION MAKING STYLE AND THE DESIGN OF DECISION AIDS: DESIGN CRITERIA

Given that we find marketing or product managers of all possible styles, how can we begin to think about designing decision aids for them? Managers of different styles tend to tackle decisions in a different way, especially in marketing because of the relatively unstructured nature of the problems and projects. On a general level, then, any decision aid will be more effective for one style than for another. It seems plausible, therefore, to suggest that for anything to be called a decision aid, it needs somehow to 'fit' the manager's style. Here, however, we run into some difficulties: (1) what does 'fit' mean, what criteria are there for such a fit, and (2) who should judge about the appropriateness of one criterion as opposed to another? Several possibilities for such criteria present themselves: a decision aid should assist the manager's superior function, or it should fill the 'gap' and assist his inferior function, or it should assist the functions that are neither superior nor inferior, or it should be uniform. These criteria undoubtedly do not exhaust the possibilities, but they give us an opportunity to begin to think in those terms. As for the question of who should be the judge of the criterion to be used, I am inclined to say that it should be the manager himself, even if that judgment can only occur indirectly, for example, by trial and error. Let us focus somewhat more in detail on each of those criteria.

The first criterion is the extent to which the decision aid assists or augments the manager's most preferred function. The notion of assisting the function is, however, not obvious at all at this point. Let us take as an example a thinking style, a person who has his thinking function most developed. We know from Jung's theory that a person of this style tends to construct formulas or models to deal with the world and wants to apply them all the time. Assisting this function may mean enhancing its capacity to build such formulas, or it may also mean to protect it from some of the dangers of an overreliance on this way of behaving. We may, for example, provide him with better tools to check the validity of his models, tools of measurement. Or we may provide him with an explicit replica of his own models so that he can work with them more freely and accurately in a quantitative sense. This was done by Clarkson, for example [6]. We notice that, for the purpose of coming up with assisting or augmenting techniques it is necessary to study how the thinking function of such a style works. Cognitive psychologists have been studying the thinking function quite extensively. In particular they have studied, somewhat in separation, four major areas: "sensation or perception (the reception and recognition of input stimuli), learning (the encoding of input information), memory (the retrieval of input information), and thinking (the manipulation of the perceived, learned and remembered information)" [18]. To the extent that each of these areas involves active manipulation of information, each involves thinking. We are mainly interested in the models the manager manipulates in his mind, and the way

in which they are modified by the perceptive functions and learning. Linkages between thinking and sensation have been touched upon by Bruner [4], and those linkages between thinking, memory and learning by Bartlett [2]. Overall, the psychological literature concerned with thinking, however, rarely goes beyond experimental work with relatively 'simple' problems. Despite that fact, some recent investigations have been concerned with the thinking of individuals in their real environment, confronted with real, relatively unstructured problems. An excellent example is the work by Elstein and his colleagues inquiring into the thinking involved in medical diagnosis by clinicians. One of their findings is that experienced physicians generate fairly specific hypotheses relatively early, well before most of the data have been collected [8]. But, take for example a person who has sensation as his most preferred function: how can his main function be enhanced? Perhaps by giving him more data, more reliable data, or more accurate or detailed data? But then, what data? And what helps the intuitive function, of the feeling function? Clearly, we have not even begun to explore these questions.

A second possible criterion is the extent to which the decision aid can fill the gap, fill in what should normally be accomplished by the inferior function. The aid can either enhance the least developed function or compensate for its inferiority. A thinking style, for example, tends to distrust 'liking or disliking' as an evaluation mode as opposed to logic. It is possible that his inferior function will, for example, get him into trouble with his salesforce. Or a decision aid can probably be designed which can enhance the inferior thinking of a feeling style, to help him analyze and organize the situation, or to weigh the evidence for various alternatives, and not the least to defend himself against opponents of his recommendations. The intuitive style will then need what we could call a sensation type aid, which assists him in focusing on details where necessary, which keeps him faced with the facts. A person with a sensation style then needs an aid that allows him to see possibilities, to look into the future, and to spark new ideas.

A third possible criterion is the extent to which the secondary and third functions are aided. This means intuition and sensation for the thinking and feeling style, and thinking and feeling for the intuition or sensation style.

A fourth possible criterion is uniformity, i.e. decision aids designed in a uniform way, adapted, as the designer may argue, to a particular kind of 'problem'. The argument of this designer contains at least two dangers, however, first of all because he designed the aid with his own style in mind, and it is normal then to find that 'his' kind of aid will not be universally accepted by all managers. The difficulty of this criterion is obvious from the literature (see for example [1, 7]). It stems from the fact that each style views the 'problem' in his particular way and handles it accordingly. Decision aid designs (for example a linear programming model, or a decision calculus model), for lack of descriptive knowledge about management and its styles, have tended to reflect the style of the designer, or at least the style of a known technique developed without such knowledge. That everybody tends to be in the grip of his style is clear from the fact that certain potential designers think that decision aids should be developed for a 'problem', not for a particular style of decision making, and that there are others, although probably a minority of current designers, who think the style is at least as important as the problem. Obviously, each style tends to define the design problem in a particular way, and it is difficult to do otherwise. The second danger is to think that an aid designed by an intuition–thinking management scientist for an intuition–thinking manager will be easily accepted by the latter. The acceptance is not at all obvious because within the style, we need to know how such a person operates in a management position. My investigations indicate that marketing managers consciously prefer that kind of aid which complements their most differentiated function, rather than an aid which helps this function itself. So this should be a major research concern for the future.

DECISION AIDS: FUNCTIONAL DESIGNS

We have seen from the preceding section that, whatever our criteria are for providing a decision maker, i.e. a marketing or product manager, with a decision aid, when it comes to designing the aid we basically get four main

types:

sensation aids
intuition aids.
thinking aids
feeling aids

This framework also lets us see that decision aids are not necessarily computerized models. These names of aids reflect their end-result in decision making and can be anything. A simulation model, for example, might serve as a thinking aid for the feeling style. A brainstorming session or synectics might perhaps be called intuition aids. 'Models' in the classic sense, such as linear programming models, decision calculus models and so on, are one particular instance of an aid. And again, whether they truly become an aid depends on the style of the potential user. A split becomes apparent between technique and end-result: different techniques may have different end-results, depending on the style of the manager. As for the design process, there are now two ways for designing and implementing a decision aid: one way starts with the technique and results in a trial and error implementation process. The other way starts with observation and description of the manager's style and the design involves a prosthesis for decision making.

SUMMARY AND CONCLUSION

The basic premise of this paper was that marketing managers tend to have different styles in the way they handle problems, projects and decisions. These style differences have consequences for communication between the manager and the management scientist, as well as for the design of the decision aid itself. In particular, a certain decision aid design will be more useful for certain managerial styles than for others. A true decision aid will need to fit the manager's style if it is going to be useful for him. Four criteria of fit were briefly explored: augmenting the manager's most developed function, assisting the least developed function, assisting the secondary and third functions, and uniformity. Regardless of what the criterion of fit is in a particular instance, we can distinguish between four general types of decision aids: thinking aids, feeling aids, sensation aids and intuition aids. The main conclusion is that before everything at this point we need research on: (1) style differences between marketing/product managers and management scientists, as a basis for communication, and (2) the characteristics of decision making of managers of different styles, as a basis for the design of decision aids.

REFERENCES

1. Ackoff RL (1967) Management misinformation systems. *Mgmt Sci.* **14**(4), B147–156.
2. Bartlett FC (1932) *Remembering.* Cambridge University Press, London.
3. Briggs-Myers I (1962) *The Myers-Briggs Type Indicator.* Consulting Psychologists Press, Palo Alto.
4. Bruner JS (1973) *Beyond the Information Given.* Norton, New York.
5. Churchman CW & Schainblatt AH (1965) The researcher and the manager: a dialectic of implementation. *Mgmt Sci.* **11**(4), B69–87.
6. Clarkson GPE (1962) *Portfolio Selection: a Simulation of Trust Investment.* Prentice-Hall, Englewood Cliffs.
7. Dearden J (1972) MIS is a mirage. *Harv. Bus. Rev.* **50**(1), 90–99.
8. Elstein AS, Kagan N, Shulman LS, Jason H & Loupe MJ (1972) Methods and theory in the study of medical inquiry. *J. Med. Educ.* **47**(2), 85–92.
9. Gell-Mann M (1971) *The Value of Schemes in Scientific Theory.* Atti del Convegno Mendeleeviano, 15–21 September 1969, Torino.
10. Hellriegel D & Slocum JW, Jr (1976) *Organizational Behavior: Contingency Views.* West, St. Paul.
11. Jacobi J (1973) *The Psychology of C. G. Jung.* Yale University Press, New York.
12. Jung CG (1971) *Psychological Types.* Princeton University Press, Princeton.
13. Kilmann RH & Mitroff II (1976) Qualitative versus quantitative analysis for management science: different forms for different psychological types. *Interfaces* **6**(2), 17–27.
14. Kolb DA (1974) On management and the learning process. In *Organizational Psychology.* (Eds. Kolb DA, Rubin IM and McIntyre JM). Prentice-Hall, Englewood Cliffs.
15. Kretschmer E (1921) *Körperbau und Character.* Springer, Berlin.
16. Leaper DJ, Gill PW, Staniland JR, Horrocks JC & de Dombal FT (1973) Clinical diagnostic process: an analysis. *Br. Med. J.* **3**, 569–574.
17. McKenney JL & Keen PGW (1974) How managers' minds work. *Harv. Bus. Rev.* **52**(3), 79–90.
18. Maier RE (1977) *Thinking and Problem-Solving.* Scott Foresman, Glenview.
19. Mason RO & Mitroff II (1973) A program of research on management information systems. *Mgmt Sci.* **19**(5), 475–487.
20. Mintzberg H (1973) *The Nature of Managerial Work.* Harper & Row, New York.
21. Singer J (1972) *Boundaries of the Soul.* Doubleday, New York.
22. Stabell CB (1974) Individual differences in managerial decision making processes: a study of conversational computer usage. Unpublished Ph.D. dissertation, MIT.
23. von Franz ML & Hillman J (1971) *Lectures on Jung's Typology.* Spring, Zürich.

ADDRESS FOR CORRESPONDENCE: *Martin De Waele, University of California, 3030 Smyth Road No. 16, Berkeley, CA 94720, USA.*

Vol. 6
1978

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 15-24, 1978. Pergamon Press. Printed in Great Britain

The Influence of Aircraft Size on Airline Operating Costs[1]

DAVID J NICOL[2]
Cleveland State University

(Received January 1977)

Many previous studies have shown that there are probably no economies of scale in the airline industry directly attributable to firm size. Scale effects within the firm, however, have often been suggested, particularly for example with regard to such firm-size independent factors as market density, route length, or aircraft size. The focus of this paper is the effect of aircraft size on an airline's unit cost. Though much of the previous evidence appears to be at variance with *a priori* expectations, the empirical results presented here suggest that most of the divergence is probably due to a conceptually inadequate basis of comparison. Appropriately compared, actual unit costs were generally found to decrease at a decreasing rate with aircraft size, though as the route length parameter is reduced some mild diseconomies appeared. Of further interest, the behavior with size of the underlying cost components was found to be quite diverse, the (far from uniform) decreases with size in unit crew, fuel, and maintenance costs being offset somewhat by increases in unit capital costs, insurance plus other costs, and landing fees.

A MAJOR issue in any price regulated industry concerns the extent to which costs are influenced by economies of scale. Usually such influences must be analyzed so that plant-level effects are distinguished from firm-level ones, but in the transportation sector further complexities arise due to the multi-dimensional nature of output. An airline, for instance, could experience economies because it handles a large number of passengers, or because it serves a larger number of routes, or because it operates a large fleet, or because it flies a large number of miles. Similarly, a firm can experience scale economies even within the context of a single route if demand on that route is high, if it operates many flights, if large aircraft can be used, or if the distance flown is long. The firm itself, of course, need not be large in order to derive benefits of the latter sort.

A number of studies have indicated that there are no firm-level economies of scale in the airline industry as long as factors of the second type are held constant. [See e.g. 2, 4, 5, 7, 8, 10, 17 and 20.] This conclusion obviously does not exclude the possibility of scale influences among these other factors, whose inherent nature are more closely associated with the characteristics of the individual route. One such influence of particular interest is the effect on costs of aircraft size, upon which the focus of this study is directed.

This would seem an essentially empirical issue, but despite the ready availability of most of the required data, it remains unresolved. On the one hand, a very reasonable *ex ante* case has been made by many for expecting continued economies of scale with increases in aircraft size. Most of the *ex post* evidence, on the other hand, does not appear to support this notion, despite the fact that most larger aircraft are also of more recent technological vintage (which presumably reinforces any scale economies). The results of this study can be viewed

[1] Based on research for a Ph.D. dissertation in the Economics Department, Case Western Reserve University. This paper represents an extended and revised version of a presentation originally given in November 1975 to the Joint National Meeting of the Operations Research Society of America and The Institute of Management Sciences (Las Vegas, NV). This earlier version appeared as a working paper (No. 73) in the Research Program in Industrial Economics at Case Western Reserve University.

[2] Assistant Professor of Finance, Cleveland State University.

as not inconsistent with either of these points of view.

Two important conclusions will become evident on reviewing these results. First, the expected decline in unit cost with increasing aircraft size is generally supported, suggesting that the divergent empirically-based findings shown elsewhere are probably due to an insufficient methodological framework. Secondly, the net decline observed in unit costs is actually the result of widely varying patterns among the various constituents. This suggests a need for a deeper probing into the behavior of each of the underlying cost components, both theoretically and empirically. This paper can also be viewed as an initial step in that direction.

The layout of the paper is as follows. First, the *a priori* case for scale economies is briefly sketched in Section A, followed by some of the apparently contradictory evidence. After describing in Section B the basic nature of an aircraft's unit cost function, in Section C estimates of the unit cost envelope curve(s) with respect to aircraft size are presented.[3] In the two following sections the behavior of each of the underlying cost components is traced, and finally in Section F, the findings are briefly summarized.

A. ECONOMIES OF SCALE: REPORTED EXPECTATIONS AND FINDINGS

The case for expecting economies in operating costs[4] with increases in aircraft size can be supported on several grounds. An aircraft, for instance, requires only one flight crew, one set of radio equipment and so on, independent of its size. Although such inputs may become somewhat larger or more expensive with size for a variety of reasons, there is a substantial basis for claiming a large measure of indivisibility in certain factors.[5] In addition, an aircraft tends to become technically more efficient as its physical size increases.[6] As a result, each of the numerous unit cost-size presentations shows unit costs declining, though at a diminishing rate, with increasing aircraft capacity—elasticities of unit cost with respect to aircraft size declining with size from approx. (minus) 30 to about 10% [see e.g. 9, 13, 15 and 16]. Continuing general acceptance of this view is evidenced by the following illustrative comments:

> "New airplanes must be bigger than predecessors in their class; [in] the drive for continued economies of scale through more seats [to] reduce the unit cost of producing a seat mile." *Aviation Week & Space Technology*, October 28, p. 42, 1974.
>
> "The scale effects in air transport economics are considerable and effective on most cost elements. In general the potential savings through the use of more productive aircraft types seems to be appreciated more highly than other scale effects and this may be due to the fact that they are easier to predict." [6, p. 137.]

Despite this belief, however, there is evidence suggesting that such expected economies have not materialized in practice.

> "Airline men say the wide-bodied aircraft never proved the big cost-savers the public was led to believe they would when they were introduced in 1970." *Wall Street Journal*, April 15, p. 29, 1975.

In addition, there are indications that the largest aircraft may not even the lowest cost. In a survey conducted by *Aviation Week*,[7] for example, the Boeing 747, the largest passenger aircraft in current use, proved to be clearly superior in unit direct operating cost only to the basic 707/DC8 types, which are less than half its size and of much older vintages; the least cost aircraft appeared to be the 'stretched' DC8's, which are about 40% larger than the original models. For the five cases reported, the 747 was more costly in three, about equal in one and least costly only once. Also unexpected was the finding that the 747's cost performance was only about equal to that of the much smaller Boeing 727 trijet in the two comparisons noted. Mixed results were reported for the two comparisons with the Douglas DC10,

[3] Aircraft size thus assumes the role of plant size in the traditional version of the long-run average cost curve.

[4] Operating costs directly attributable to particular aircraft are called direct, the most important of which include flight crew costs, fuel costs, aircraft maintenance expenses, and associated capital expenses. All other operating costs (e.g. passenger service expenses, advertising, general and administrative expenses, etc.) are treated as essentially independent of aircraft type and called indirect. Operating costs are roughly evenly divided between the two types.

[5] Douglas and Miller [4], for example, state that (p. 10): "...crew requirements [are] less than proportional [to aircraft size]" and "...there are many equipment costs that are independent of aircraft size,...and others which are not strictly proportional...."

[6] "The general case for the larger vehicle is reinforced in the transport aircraft by its improved aerodynamic and propulsive efficiency with increasing size and Reynold's number..." Smelt [15, p. 513].

[7] Each of fourteen airlines was requested to compare its experience with the 747 to each of the other types it operates; see *Aviation Week & Space Technology*, March 27, pp. 31–34, 1972.

but this ambiguity might be expected since the somewhat smaller DC10 is also slightly newer than the 747.

Of further interest is the wide range in the reported differentials of unit direct operating costs, even in the one set of comparisons where the 747 appeared clearly superior. The unit cost of the 747 varied from 12 to 43% lower than the basic 707/DC8 type (in one of the fourteen comparisons it was actually higher). Ambiguities concerning the magnitude of such relative cost margins is also suggested by the recent Civil Aeronautics Board (CAB) decision to investigate National Airlines' proposed "No Frills" fare:

> "...partly because it doubts National's claim that perseat operating costs of wide-bodied planes are 36% below those of conventional craft..." *Wall Street Journal*, April 15, p. 29, 1975.

B. UNIT COST FUNCTIONS

Because transportation output measures encompass multiple dimensions, the relationship of costs to output must be carefully considered. In air transportation, the most commonly used measure is available seat-miles (S), which can be viewed as the product of three separate factors:

$$S = C \cdot R \cdot D,$$

with C representing the average aircraft's number of seats; R the average miles per flight (usually called stage length); and D the total number of departures for the period. For a given aircraft type and seating configuration, C can be treated as a constant. Similarly, by limiting cost comparisons to a particular stage length, R can be held constant—i.e. treated as a parameter. With these restrictions, S is then uniquely related to D, which allows the number of departures to be viewed as the basic determinant of both output and costs. This variable can assume any value between zero (the aircraft can be grounded) and some technologically determined maximum (at most a daily limit determined by 24 hr minus the necessary time for turnarounds).

Total direct operating costs for a given aircraft are often split into capital costs (i.e. the explicit annual lease cost or implicit opportunity cost) and all other costs, usually referred to as flight costs. At it turns out, so long as stage length is fixed, flight costs per departure can be treated as a constant which increases less than proportionately with stage length. Thus the unit cost function for an aircraft can be seen to be of very simple form:

$$\frac{DOC}{S} = \frac{F(R)}{C \cdot R} + \frac{L}{C \cdot R} \cdot \frac{1}{D},$$

where DOC is the aircraft's direct operating cost (e.g. per year); $F(R)$ the flight costs per departure; L the (e.g. annual) capital cost; and S, C, R, and D are as previously defined. With stage length fixed, the unit cost function of a particular aircraft is thus just the sum of a constant, representing unit flight costs, and an amount inversely proportional to output, representing the capital cost allocation; i.e. minimum unit DOC is determined by maximum departures. The justification for retaining stage length as an explicit parameter should be evident.

C. THE UNIT COST ENVELOPE

Unit cost functions as described above were developed for each operational aircraft during 1972.[8] The resulting cost-curve envelopes are shown in Fig. 1 for five representative stage lengths. For clarity, unit costs are shown here in dollars per seat (which for a given stage length readily converts to cents per seat-mile), and output is shown as daily seats available (converted from annual seat-miles). The results clearly verify the existence of economies due to aircraft size. Additionally, as stage length is reduced, some mild diseconomies can be seen to emerge for the largest aircraft.[9]

[8] These estimates are based upon data reported by the airlines to the Civil Aeronautics Board (CAB) and published in their annual *Aircraft Operating Cost and Performance Report* [18]. The 1972 results, published in July 1973, were chosen to approximate the pre-'fuel crisis' equilibrium. Further details on these estimates and the procedures followed are contained in Nicol [12], Chapters III through VIII.

[9] In other words aircraft size for minimum unit DOC increases with stage length. For stage lengths below about 300 miles the 727-200 appears least costly, between about 300 and 1300 miles, the DC8-63 is lowest, and above about 1300 miles, the 747 is best. Interestingly of the seven aircraft needed to describe the cost envelope(s) when stage length is fixed, only three are necessary to describe the envelope of minimum unit DOC with respect to stage length.

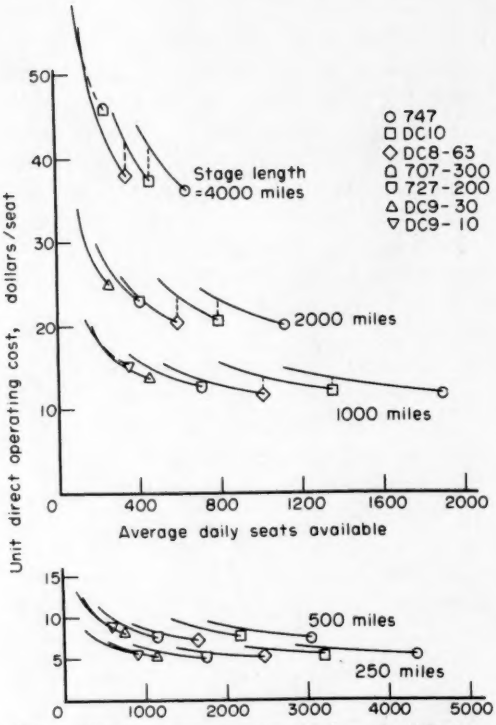


FIG. 1. *Unit direct operating cost envelopes for* 1972 *domestic operations.*

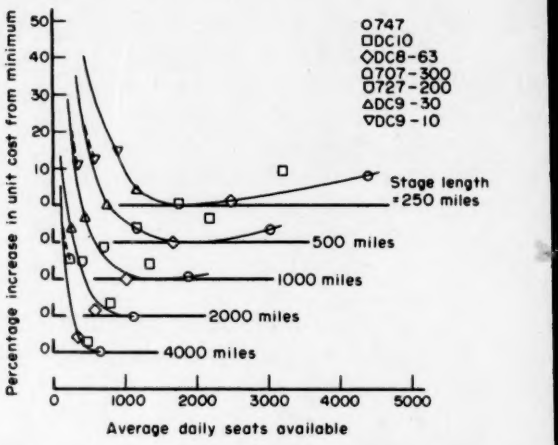


FIG. 2. *Relative unit direct operating cost envelopes for* 1972 *domestic operations.*

To give a clearer perspective on the relative effects on unit costs, Fig. 2 shows the cost curve envelopes expressed as percentage increases from the minimum for each of the five stage lengths. At output levels about half that for which the cost minima occur, for instance, units costs are about 10% above the minimum, implying an elasticity of unit cost with respect to output of about 10% (i.e. a 10% unit cost reduction can be associated with a 100% increase in output).[10] The diseconomies that eventually do occur at the shorter stage lengths are seen to be not too significant. Further, to focus more closely on the influence of aircraft size, Fig. 3 shows the same information with respect to aircraft capacity (rather than output).[11] It is also apparent here that 'plane-stretching' has a more dramatic cost effect than size *per se*. Comparing the DC9-10 to its stretched counterpart, the DC9-30, and similarly the 707-300 to the DC8-63, for example, shows elasticities of about 40 or 50%. Stretching thus seems to cause a unit cost reduction on the order of about half the percentage increase in seating capacity.

D. UNIT CAPITAL COSTS

Of further interest is the variation with aircraft size of the unit cost components underly-

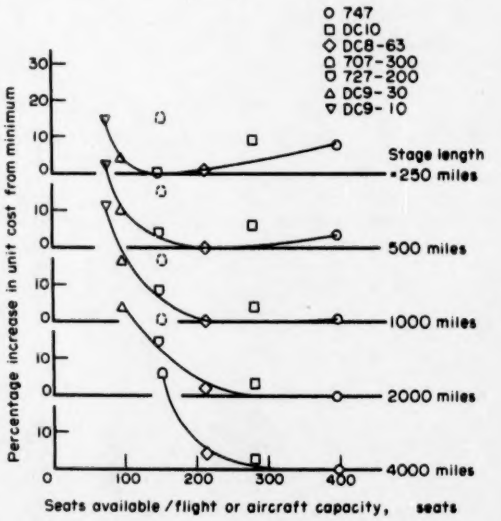


FIG. 3. *The influence of aircraft capacity on relative unit direct operating cost minima in* 1972 *domestic operations.*

[10] Figure 2 may also serve as a point of departure in quantifying the influence of market density on costs. For instance, a market capable of supporting only DC9-10's inherently has about 15% higher unit *DOC* than one sufficient to support 727-200's. Since demand is not in general independent of flight frequency and aircraft size, however, considerable further elaboration of this cost information would be necessary to derive the equivalent of a supply curve. See, for example, Straszheim [17], Miller [8], Douglas and Miller [3, 4].

[11] Following Douglas and Miller [4, p. 24] a 100% coach equivalent seating configuration was used throughout, with a seat width of 18 in. and pitch of 34 in.

ing the above totals, particularly since it turns out that such relationships are far from uniform. One of the largest cost factors is capital cost, whose importance in unit *DOC* rises considerably as aircraft size is increased. Capital costs as a proportion of minimum unit *DOC* (i.e. for maximum utilization) range rather uniformly with size from just over 20% of *DOC* for the smallest aircraft to almost 40% for the largest.[12]

Unit capital costs can be viewed as the product of three components, recalling that $S = C \cdot R \cdot D$:

$$\frac{L}{S} = \frac{L}{C} \cdot \frac{1}{R} \cdot \frac{1}{D}.$$

Since stage length must be held constant in any comparison, the influence of aircraft size will be exerted either in the first term (capital costs per seat) or the third (the inverse of annual departures). Figure 4 shows that maximum departures vary only slightly with aircraft type, stage length having a much greater influence.[13] This suggests that our attention here can be directed to the effect of size on annual capital costs per seat of aircraft capacity (or equivalently, investment costs per seat), estimates of which are presented in Fig. 5.[14] The upward trend in capital cost per seat with increasing size is evident; ranging here from about $6000 up to about $8000 per year per seat.

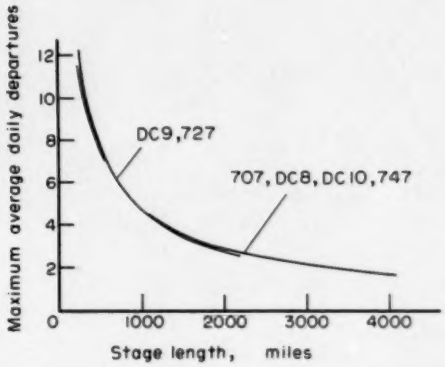


FIG. 4. *The effect of stage length on maximum average daily departures in 1972 domestic operations.*

E. UNIT FLIGHT COSTS

The effect of size on unit flight cost is of course thereby more pronounced than it was for total unit *DOC*, since some of the decrease in unit flight cost with aircraft size is offset by the increase in unit capital costs.[15] Basically, unit flight costs steadily diminish at a decreasing rate until about the level of the DC10, beyond which no further decreases are evident (the 747 appears to be very slightly higher).[16] Also of interest, the influence of air-

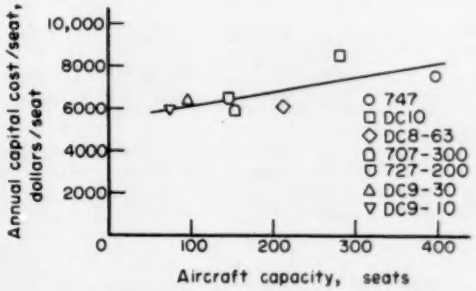


FIG. 5. *The influence of aircraft size on annual capital costs per seat, 1972.*

[12] The effect of stage length on these proportions is relatively unimportant. Of course at less than full utilization these figures will increase.

[13] These estimates were derived from a scheduling model based on observed maxima with respect to stage length for each aircraft type and thus form the basis for each aircraft's estimated maximum practical seat-mile capacity.

[14] Annual capital costs were determined in two stages. First, a 1972 price was estimated for a fully equipped aircraft of each type (with spares) from information reported in *Aviation Week and Space Technology*, *The Wall Street Journal* and various other sources. The resultant estimates were quite similar to those of Simpson [14]. Second, a ratio of annual lease charges to price (i.e. investment) was estimated for 1972 from the CAB data [18]. This could be derived only in a few cases because of data limitations, but these were observed to differ only slightly from an average ratio calculated by neglecting the year of acquisition (whether leased or purchased). For each type then, the average ratio was used to approximate the relationships prevailing during 1972. Furthermore, since this ratio varied little among types, ranging from about 0.10 to 0.11, the average of 0.104 was used. This ratio times the 1972 price yields the desired estimate of each aircraft's annual cost of leasing or opportunity cost of owning.

[15] As mentioned earlier, as long as stage length is fixed flight costs per departure can be considered constant, i.e. unit flight costs are also constant given stage length.

[16] A note of caution is in order here concerning the implications of Fig. 1 regarding the economies of scale (as opposed to size) question. The DC10, first used in 1971, is technologically the newest aircraft. Based upon year of first commercial application, the 747 is about a year older, the three stretched aircraft (the DC9-30, 727-200, and DC8-63) about four years older, the DC9-10 about 6 years older, and the original (turbofan) 707-300 types about 8 or 9 years older. The 'economies of scale' curves for 1972 (i.e. with DC10 technology) would probably be flatter than those of Fig. 1, and might possibly show continued decreases with size beyond the 747.

craft size gets progressively stronger as stage length increases, particularly at the longer stage lengths where only the relatively expensive 707-300 type is available among the smaller aircraft.

A further breakdown of total flight cost is necessary since each of its components behaves differently with aircraft size. Three major categories—maintenance, crew, and fuel—are of about equal importance, and together they account for 90% or more of flight costs. Their behavior with size, however, differs sharply. Thus, while maintenance represents an essentially constant 32% of flight costs for all aircrafts, the crew proportion decreases from about 38 to 24% as size goes from smallest to largest, and the fuel proportion increases with size from about 26 to about 35% (both rather uniformly). The flight cost proportion for the two smallest components—insurance plus other costs and landing fees[17]—also increases with size from about 4 to about 9%.

These five components can be usefully split into two functional groupings. Three of the five—crew, maintenance, and insurance plus other costs—can perhaps best be described as 'block-hour dependent'—meaning that they depend on ramp-to-ramp time, i.e. on the total time elapsed from the moment the aircraft first moves under its own power (on departure) until it finally comes to rest after landing (on arrival). The other two—fuel and landing fees—can be described as 'departure-dependent'.[18]

For each of the first three, the cost per departure (i.e. the $F(R)$ of page 17) can be expressed as:

$$F(R) = f \cdot h(R),$$

where f is the particular input's cost per block-hour, and h (R) is the block-hours per flight, both of which in general depend on aircraft type. The cost per seat-mile for each of these three components then becomes:

$$\frac{F(R)}{C \cdot R} = \frac{f \cdot h(R)}{C \cdot R} = \frac{f/C}{V(R)},$$

where $V(R) = R/h(R)$ is block speed in miles per block-hour, and the constant f/C is a cost per hour per seat.

The stage length influence on these unit costs is directly traceable to its effect on block speed. The dependence of block speed on stage length is shown in Fig. 6[19]—note in particular the sharp drop in block speed below about 1500 miles. Of further interest here, however, is the observation that for any given stage length, block speed is relatively unaffected by aircraft type.

This last finding suggests that the essential factor through which aircraft size influences these unit costs is f/C, i.e. the input cost per block-hour per seat of aircraft capacity. Estimates of the latter are presented in Fig. 7 with respect to aircraft size for each of the three hourly components individually, and for their total.[20] Hourly crew cost per seat can be observed to consistently decline with increasing aircraft size (despite a uniform increase in the crew cost-per-hour coefficient). The unit crew cost of the largest aircraft, for instance, is only about half that of the smallest ones; even the

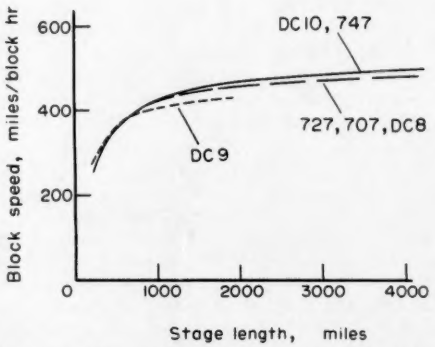


FIG. 6. *The effect of stage length on estimated block speed in 1972 domestic operations.*

[17] Landing fees are not included in *DOC* by the CAB, but rather they are treated as indirect and considered as part of aircraft and traffic servicing expenses. Because of their dependence on aircraft type, however, they have been included here.

[18] Actually, from an estimation standpoint, this split can be viewed as unnecessary since, for a given aircraft on a particular route, the independent variables in each component cost function will differ only by the multiplicative constant of block-hours per departure (cf. the next footnote). For present purposes, however, the distinction is desirable to maintain a correspondence to normal input and price measures (e.g. crew-hours and price per crew-hour).

[19] These were readily obtained from the 1972 CAB data [18] via simple regression with the following generally accepted functional form: $h(R) = t_0 + t_1 \cdot R$ where t_0 and t_1 are constants (e.g. see Douglas and Miller [4, pp. 20–21]).

[20] These are based on the 1972 CAB data [18]. In light of the probable existence of inter-firm cost level differences (e.g. see Boske [1]), considerable effort was necessary here to avoid biased coefficients, particularly since not all aircraft types were operated by every firm.

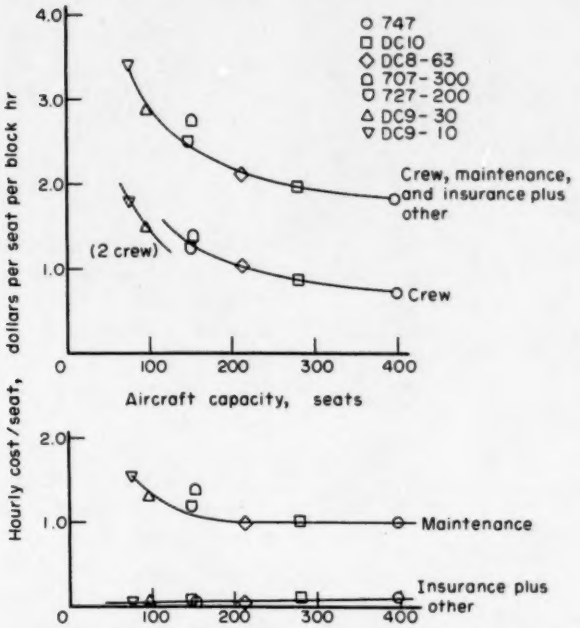


FIG. 7. *The influence of aircraft size on hourly costs per seat for crew, maintenance and insurance plus other costs, 1972 domestic operations.*

use of only two crew members in the DC9's is insufficient to offset their relatively small number of seats. Similarly, unit maintenance costs show a considerable initial decrease with size up to about 200 seats, approximating only about two-thirds the costs of the smallest aircraft. Beyond about 200 seats, however, no further advances due to size are in evidence. Finally, the insurance plus other component, though relatively small, actually increases on a unit basis by about 50% as size goes from smallest to largest.

Consider now the two departure-dependent costs. The first is fuel cost per departure, which is essentially just the product of price per gallon of fuel and gallons of fuel required per departure:

$$F(R) = P \cdot G(R, W),$$

where P is a net price, and $G(R, W)$ in gallons is dependent upon the flight length and payload weight (passenger and freight) as well as aircraft type. Unit fuel costs can thus be expressed as:

$$\frac{F(R)}{C \cdot R} = \frac{P}{C \cdot R/G(R, W)}$$

i.e. as the price per unit of fuel divided by fuel productivity (in seat-miles per gallon). Since price is essentially constant by type,[21] variations in unit fuel cost with size arise through fuel productivity differences.

Estimated unit fuel costs and corresponding fuel productivities for each type during 1972 are shown in Fig. 8 as a function of stage length.[22] As with block speed, the stage length effect is quite important, particularly at the shorter ranges. Fuel productivity, however, is also very dependent upon aircraft type, though the picture with respect to size is rather mixed. The newest aircraft (the DC10) generally appears to have the highest fuel productivity, the oldest (represented by the 707-300) has the lowest, and the others basically rank in accordance with their size (somewhat excepting the DC9-30).

A clearer picture emerges when the fuel re-

[21] The average 1972 domestic price of 11.5 cents per gallon was adjusted upward slightly according to aircraft type by between about 2 and 3½ percent to reflect oil expenses and related taxes.

[22] Since fuel requirements are relatively insensitive to payload within the variations normally observed, W was held constant for each type by assuming a passenger load of one-half capacity and a weight per passenger of 227.8 pounds, the 1972 weighted average including freight. Also, the DC10 estimate is shown dashed beyond about 1500 miles because this represents an extrapolation based upon a limited number of observations.

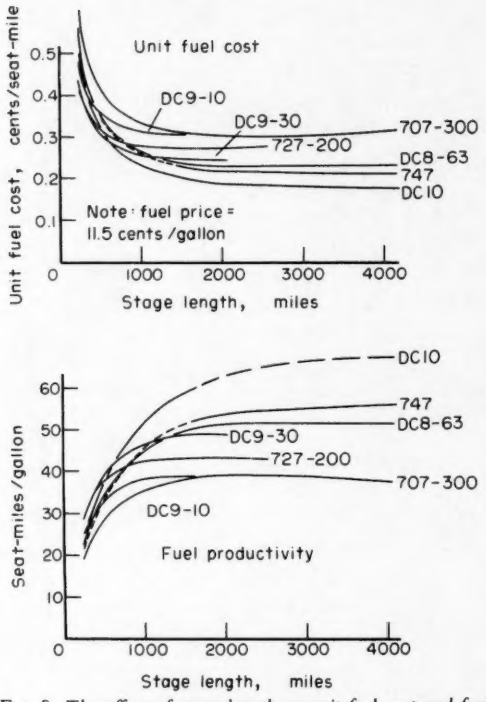


FIG. 8. *The effect of stage length on unit fuel cost and fuel productivity, 1972 domestic operations.*

quirements function is broken into its component parts:

$$G(R, W) = g(R) \cdot W$$

where W is the 'gross payload' or 'landed weight' in pounds (defined as take-off gross weight less the fuel used for the flight), and $g(R)$ is the estimated function for fuel required per pound of gross payload.[23] Fuel productivity thus becomes the product of two terms:

$$\frac{C \cdot R}{G(R, W)} = \frac{R}{g(R)} \cdot \frac{C}{W},$$

where $R/g(R)$ is pound-miles of gross payload per gallon, which for a given aircraft type depends only on stage length; and C/W is each aircraft's seating capacity per pound of gross payload, which depends essentially only on aircraft type (though slightly affected by actual payload).

The first term, which can be viewed as representing the underlying 'technical efficiency' of each airframe/engine combination, is shown in the upper portion of Fig. 9 related to stage length. Excepting the DC10, which appears superior to all others, this productivity measure increases rather uniformly with aircraft size, at least partially confirming *a priori* expectations.[24] The second term, which can be looked on as a 'structural efficiency', is shown

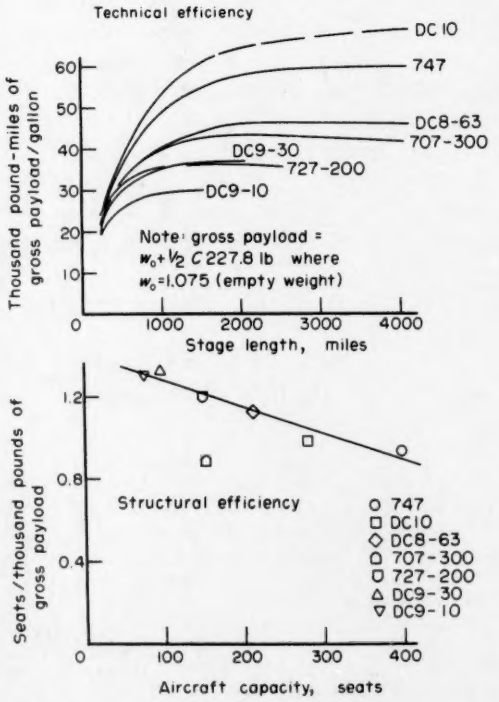


FIG. 9. *Fuel productivity: the effect of stage length on 'technical efficiency' and the relationship of aircraft size to 'structural efficiency'.*

[23] Following a development similar to that used by Mutch [11], the fuel requirements function can be shown to be of the approximate form: $G'/(W + 1/2G') = A(B + R)$, where G' is $G(R, W)$ converted to pounds, and A and B are constants (note that $W + 1/2G'$ represents an average aircraft weight during the flight). The CAB data [18] yields observation sets on G', R, and passenger plus freight weight, and W was approximated as the sum of this last item and a constant (published operating weight empty figures were adjusted upwards by 7.5% to reflect fuel reserves, crew weight, etc.). The A and B coefficients were then directly estimated using ordinary least-squares on the ratio of fuel to average aircraft weight and stage length. The resulting estimate of the fuel ratio function can be readily rearranged to give $g(R) = G/W$, which depends only on stage length and aircraft type.

[24] Note that the individual engines of all but the two largest aircraft are physically quite similar, and that the physical size of each aircraft roughly corresponds to its seating capacity (though the 727-200 is physically only slightly larger than the DC9-30). The DC8-63 is basically an aerodynamically improved (as well as stretched) 707/DC8. The 747 is substantially larger than the others and has more advanced and much larger engines. While the DC10's physical size is between the 747 and the stretched DC8, it also has much larger engines, and these are of the latest technology.

Vol. 6
1978

in the lower portion of Fig. 9 as a function of seating capacity (remember that W includes a 50% passenger load). For this productivity measure, a downward trend with size is evident.[25] Thus the general decline in unit fuel cost with increasing size can be seen to be the net result of two tendencies working in opposite directions as aircraft size increases.

The structural efficiency measure can also be used to describe the increase with size in landing fees per seat-mile—the last flight cost component. Landing fees per departure are determined by a charge per pound landed weight, so that its $F(R)$ can be expressed as:

$$F(R) = k \cdot W,$$

where k is an average cost per pound and W is as previously defined.[26] Landing fees per seat-mile thus become:

$$\frac{F(R)}{C \cdot R} = \frac{k/R}{C/W},$$

which for a given aircraft decline proportionately with stage length. Holding stage length constant, and since k is constant, the behavior of unit landing fees with size thus directly depends on C/W. This productivity measure has already been shown in Fig. 9 to decrease with size, which accounts for the increase with size in unit landing fees.

F. SUMMARY OF FINDINGS

The results presented here show that unit *DOC* tends to decline with increasing aircraft size, and thus are not in conflict with *a priori* notions concerning economies of scale in *DOC*, though there is some indication that the magnitude of this effect may be less than usually estimated. Interestingly, however, these results are also not in conflict with the somewhat contradictory *ex post* observations noted earlier.

This analysis has emphasized the importance of the stage length parameter in determining the level of the cost function, and the importance of utilization (e.g. annual departures per aircraft) in determining output and thus cost. Any cost comparison neglecting to account for either of these factors will in general not be valid. Furthermore, though unit costs generally do tend to decline with increasing aircraft size, some diseconomies with size are in evidence at the shorter stage lengths. Thus, the framework and results presented here, summarized in Fig. 1, provide a means to explain both the wide variation in reported cost differentials among aircraft types, as well as the occasionally inconsistent cost rankings, without requiring a major revision of existing theoretical expectations.

Of further interest and importance, however, is the finding that the underlying components of *DOC* do not tend to behave uniformly with size. Unit capital costs, insurance plus other costs, and landing fees actually increased with aircraft size, while unit crew, fuel, and maintenance costs essentially decreased. Furthermore, the influence of aircraft size was diverse even among these last three items, as unit crew costs decreased rather steadily from the smallest aircraft to the largest, unit maintenance costs flattened out about midway through the size range, and the relationship of unit fuel cost to aircraft size was actually quite mixed (particularly at the shorter stage lengths), the net result depending on two opposing influences.

[25] Some of this decrease must be attributed to the tendency for design range to increase with size, causing the aircraft to be heavier, *ceteris paribus*.

[26] A cost per thousand pounds landed weight of 41.3 cents was used for 1972, which corresponds to the Air Transport Association's estimate that these averaged 45 cents in 1973, about 9% higher (see *Aviation Week and Space Technology*, May 27, 1974, p. 27). Essentially the same result was obtained by relating the landing fee per departure data for 1969 from CAB Docket 28166-9 [19] to landed weight, and inflating this figure to 1972 (using the average growth rate in landing fees over the 1967–73 period, as reported by the Air Transport Association in the article above).

REFERENCES

1. Boske LB (1976) Cost efficiency in the domestic trunk airline industry. *J. Econ. Bus.* **28**, 128–137.
2. Caves RE (1962) *Air Transport and Its Regulators.* Harvard University Press, Cambridge.
3. Douglas GW & Miller JC, III (1974) Quality competition, industry equilibrium and efficiency in the price constrained airline market. *Am. Econ. Rev.* **LXIV**, 657–669.
4. Douglas GW & Miller JC III (1974) *Economic Regulation of Domestic Air Transport: Theory and Policy.* Studies in the Regulation of Economic Activity, Brookings Institution, Washington, DC.
5. Eads GC, Nerlove M & Raduchel W (1969) A long-run cost function for the local service airline industry. *Rev. Econ. Stats* **51**, 258–270.
6. Elle BJ (1968) The economics of convenient airline service. In *Passenger Transportation*, Michigan State University.

7. JORDAN WA (1970) *Airline Regulation in America: Effects and Imperfections*. John Hopkins Press, Baltimore.
8. MILLER JC III (1970) Testimony CAB Docket 21866-7, Exhibit DOT-T-1, Washington D.C.
9. MILLER RH (1971) Some Air Transportation Concepts for the Future. *J. R. Aeronaut. Soc.* **75**, 431–455.
10. MURPHY NB (1969) Sources of Productivity Increases in the US Passenger Airline Industry. *Transp. Sci.* **3**, 233–238.
11. MUTCH J J (1973) The potential for energy conservation in commercial air transport. RAND Corporation Report R-1360-NSF.
12. NICOL DJ (1977) Economies of scale in the production of air transportation. Ph.D. Dissertation, Department of Economics, Case Western Reserve University.
13. PEARSON H (1971) The development of propulsion systems for air transport. In *Air Transport*. Air Transport Assoc. of America, Washington, DC.
14. SIMPSON RW (1972) Technology for design of transport aircraft. In *Proceedings of the NASA/MIT Workshop on Airline Systems Analysis*, MIT, Flight Transportation Laboratory.
15. SMELT R (1971) Looking ahead in aeronautics and astronautics—a U.S. View. In *The Future of Aeronautics* (ALLEN JE & BRUCE J, Eds), St Martin's Press, New York.
16. STRATFORD AH (1973) *Air Transport Economics in the Supersonic Era*. Macmillan, London.
17. STRASZHEIM MR (1969) *The International Airline Industry*. Brookings Institution, Washington, DC.
18. U.S. CIVIL AERONAUTICS BOARD (1973) *Aircraft Operating Cost and Performance Report*. Government Printing Office, Washington, DC.
19. U.S. CIVIL AERONAUTICS BOARD (1970) Bureau of Economics, CAB Docket 21866-9, Washington, DC.
20. U.S. CIVIL AERONAUTICS BOARD (1975) *Report of the CAB Special Staff on Regulatory Reform*. Washington DC.

ADDRESS FOR CORRESPONDENCE: *Dr DJ Nicol, Cleveland State University, College of Business Administration, Cleveland, OH 44115, USA.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 25–31, 1978. Pergamon Press. Printed in Great Britain

Step-Through Simulation[1]

JACOB W ULVILA

Harvard University

REX V BROWN

Decisions and Designs Inc., Virginia, USA

(Received February 1977; in revised form July 1977)

This report describes the initial development of the Step-Through variant of Monte Carlo simulation, a new procedure for implementing decision analysis or for training decision makers. Like regular Monte Carlo simulation, it involves sampling possible aftermaths of an initial action, and generating a distribution of outcome values for it. However, the detailed structure and/or assessments of the decision model are elicited *as called for* in the execution of each trial. It, therefore, permits substantial economy of elicitation if there are few trials. The Step-Through procedure also offers economy of elicitation and calculation over a traditional extensive tree decision-analytic model without requiring simplifications or aggregations in the model's conceptualization. In addition to describing this procedure, this paper presents the results of a preliminary test and evaluation of its viability.

1. MOTIVATION FOR STEP-THROUGH SIMULATION

IN THE standard, classical use of decision analysis to model a decision maker's action selection problem, the 'extensive tree,' relevant elements of the situation are modeled explicitly in a decision tree. That is, the decision maker's initial options and their subsequent event and action implications are displayed in a decision tree, a probability is assessed for the occurrence of each event, and a value is assigned to each path through the tree. On the basis of these assessments, each option can be evaluated by taking a weighted average value, by the traditional method of 'rolling back' the tree [3]. (It should be noted that there are many acceptable alternative formulations of the decision model within this mold, differing, for example, in how finely the events are defined and therefore how complex the tree is. For convenience, we will reserve the term 'extensive tree' for the more complex model formulations.)

This use of decision analysis is conceptually very appealing, as it allows the maximum amount of disaggregation of the problem into manageable sub-problems. However, in practice, an explicit model of all relevant elements of a decision problem often requires an unmanageably large and complex assessment task, sometimes involving millions of assessments. In addition, the computational demands of a complex extensive tree are very large.

In an attempt to overcome the computation and elicitation burdens of the complex extensive tree model, some decision analysts have sought variants of this classical paradigm. Some of these variants include: direct-value model, direct-probability model, and staged-tree model [2].

The direct-value model is a very simple one that models the value of the initial decision options explicitly but considers uncertainties and subsequent action choices only implicitly.

[1] The research and development effort reported in this paper was sponsored by Office of Naval Research Contracts N00014-74-C-0263 and N00014-75-C-0426 and Rome Air Development Center Contract F30602-74-C-0225.

In a direct-value model, each option is evaluated along one or more dimensions of value, and a trade-off function is specified to allow the resulting values to be aggregated into a single index of attractiveness. Uncertainties and subsequent actions are taken into account implicitly in deciding the values of each immediate option. The direct-value model greatly reduces the amount of both elicitation and computation required by an extensive tree, at a cost, however, for the direct-value model often requires assessments that are very difficult to make. That is, by requiring evaluations that implicitly consider the uncertainties and subsequent actions, the direct-value model aggregates the entire path following an option. This procedure works well for fairly simple decision problems in which the consideration of value trade-offs is crucial but there are no critical uncertainties. However, in complex decision situations, the aggregation required in this procedure may lead to assessments that are grossly inaccurate because the implicit considerations are so difficult to make. The problem is that everything that might occur after the action must somehow be reflected in the values assigned to the action.

Similar to the direct-value model in its simplicity is the direct-probability model, where probabilities are assessed explicitly for key uncertain events that directly contribute to the outcome's value. This variant only implicitly considers multiple value criteria, subsequent action choices, and events that provide information rather than contribute directly to value. A direct probability model works well in simple decision situations in which uncertainty, rather than value, is the main consideration and determinant of a good decision. However, in complex decision situations, the direct-probability model, like the direct-value model, does not allow for the amount of disaggregation required to make accurate input assessments.

The staged-tree variant is representative of a class of variants that force the decision problem to take a special form. The staged-tree variant assumes that all of the relevant information in an extensive tree model of a decision problem can be represented in a series of summary state descriptions. The decision tree can then be replaced by a concatenation of Markov models. This special form greatly reduces the complexity and computational burden of the extensive tree but at the expense of introducing stochastic independence assumptions that are not easily met in practice. In addition, the staged-tree and other related variants do not permit the convenient disaggregation that is permitted in the extensive tree.

All of the simplifications of the extensive tree mentioned above are effective in reducing the number of elicitations and the amount of calculation needed in a large model. However, none of these methods maintains a very desirable feature of extensive tree, that of disaggregating a complex decision problem into a series of sub-problems that are easy to manage. One way to reduce the computational burden of the extensive tree analysis while maintaining its disaggregation is to use conventional Monte Carlo simulation to sample paths through a tree that has been completely specified to obtain a distribution of end positions for each immediate action alternative. This method allows for the same degree of disaggregation as the extensive tree, but it also requires the same number and types of elicitations. Thus, a gain is made only in the amount of computation required; the assessment burden is the same as that for an extensive tree.

The problem remains to find a modeling technique that allows for a reduction in the number of elicitations and the amount of calculation required by a complex extensive tree analysis but retains the disaggregation advantage that an extensive tree offers for complex decision situations. The **Step-Through** variant of Monte Carlo simulation appears to have these features.

In addition to providing a complete analysis of alternative options leading to a clear prescription for action, a desirable analytic tool may have other functions such as:

1. To organize and sensitize a decision maker's perceptions of a decision problem so that he can more readily make an informal decision or provide inputs to some other decision model. For instance, a decision maker may be better prepared to make the aggregated assessment of summary outcomes needed for a direct probability model if he first explores the possible sequels to an initial decision in a detailed structured way.
2. To train or evaluate a tactical decision

maker by giving him vicarious experience in responding to a rapid succession of contingencies flowing from some initial action. War gaming has exactly this function, but its conventional forms often lack a satisfactory probabilistic mechanism.

Again, Step-Through simulation would appear to be useful for these functions as well as for analyzing complete decision problems.

2. DESCRIPTION OF STEP-THROUGH SIMULATION

The logic of Step-Through simulation is essentially that of conventional Monte Carlo simulation. That is, Step-Through involves simulating paths through an implicit extensive tree in order to obtain a sample of endpoints for each initial decision option. Unlike conventional Monte Carlo simulation, however, Step-Through requires assessments of inputs only *as a trial progresses*; it uses an interactive procedure that alternates assessments with Monte Carlo sampling of events in chronological sequence, rather than a complete specification of all inputs in advance. In this manner, many of the tree paths that have a low probability of occurrence will never be sampled and thus will never need to be modeled.

A process diagram of the Step-Through procedure is shown in Fig. 1. The procedure is begun by the decision maker's choosing an option from among those immediately avail-

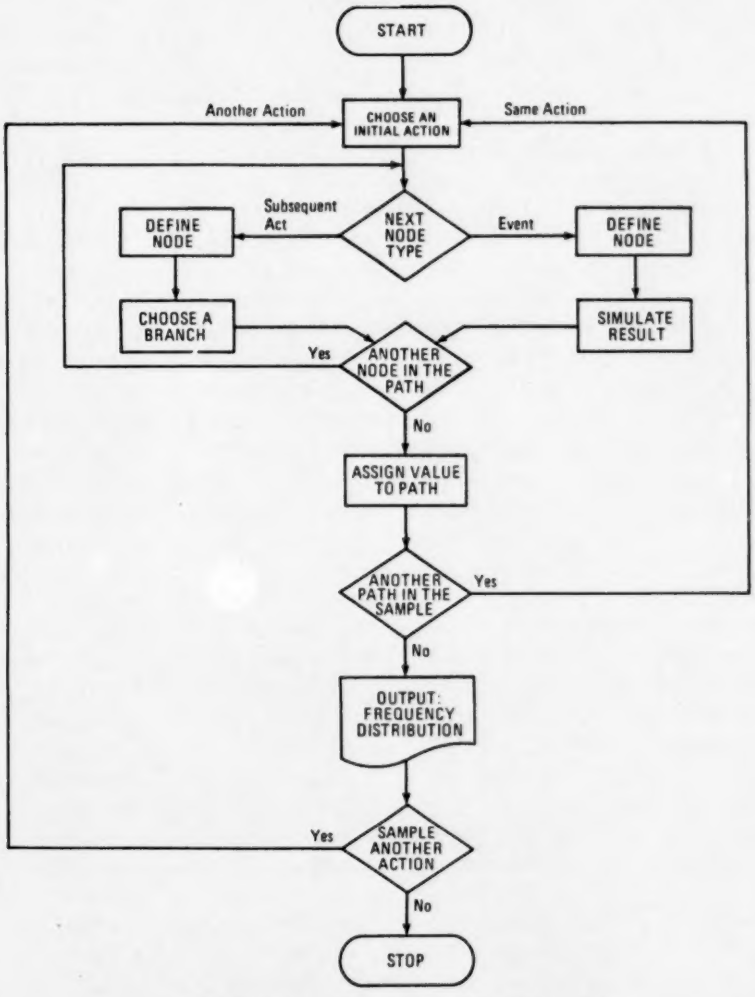


FIG. 1. *Process diagram of Step-Through simulation.*

able. Next, the following node in an implicit decision tree is defined. If this node is an event node, Monte Carlo simulation is used to choose a branch of the node; this simulates a result at that event node. If this node is an action choice node, however, the decision maker chooses the branch to follow. Nodes are processed in this manner until a complete path through the tree has been simulated, at which time a value for the trial path is assigned. Then, other paths that follow the same initial action are simulated, in a like manner, generating a sample of endpoints for that initial action. The same procedure is then repeated for each initial action option. After paths have been sampled for all initial options and their endpoints valued, their frequency distributions of values are compared to determine the best choice. (The valuation may be accomplished by assigning a single value to each path, or it may require assignments along several dimensions of value.)

This full-scale form of Step-Through can be used to analyze a complete decision problem and the resulting samples of endpoint values are interpretable in the same way as those for conventional Monte Carlo simulation. However, Step-Through offers a reduction of effort and a flexibility of operation that is not available in conventional simulation. The total number of required assessments is fewer (since assessments are required for only the paths actually encountered in this simulation), and the structure of the tree itself can be developed as the trial progresses. Furthermore, at the user's option, subsequent acts can be chosen consciously as they occur, specified as elements of a predetermined strategy, or treated as uncertain events [1].

In addition to this full-scale form of Step-Through, a special abbreviated form, which can be used as a preparation for a larger analysis, is **Modal Step-Through.** In Modal Step-Through, a single trial is generated for each initial action. However, instead of eliciting a probability distribution for each uncertain event encountered and then randomly sampling an outcome from it, the most probable, or modal, outcome is selected judgmentally. This procedure will not, of course, provide any probabilistic information about the total value of an initial action, but it is very easy and inexpensive to do and it may suggest the best way to proceed with a larger analysis. Modal Step-Through may also be very useful in training, or at least preparing, a subject for the elicitations that he may be called upon to provide. For example, this form has occasionally been used to precede some direct-probability models of decision problems. These models appeared to benefit from having had the subject address, however briefly, the detailed texture of possible aftermaths to the initial actions through the Model Step-Through procedure.

2.1 *The number of trials needed*

As with other kinds of Monte Carlo simulation, the number of trials in a Step-Through simulation must be sufficiently large for the sample to adequately resemble the implicit population of paths from which it is drawn. However, the issue of what constitutes a sample that is 'large enough' is particularly crucial with Step-Through because of its exceptionally high variable cost per trial. In particular, sampling more than a few trials will soon offset Step-Through's savings in fixed cost compared with conventional Monte Carlo simulation.

One might think that the sampling error in the estimates of the population's statistics would be related to the length of the path, and it might therefore be inferred that using Step-Through to model a decision problem with long paths requires an unreasonably large sample of paths. A closer examination, however, shows this concern to be unwarranted.

Since Step-Through is a procedure for drawing a random sample from a target population (in this case, the target population is the value of endpoints in a decision tree), the rules of error for sample statistics apply. These rules, however, state that the sampling errors of the statistics depend only on the type of sample (for instance, a random sample, a stratified sample, a systematic sample, and so forth), the sample size, and the underlying distribution of the target population. The sampling errors do not depend upon the complexity of the model used to represent the target population. For example, in the language of classical inference, the standard error of a random sample from a normally distributed population is equal to the standard deviation of the population divided by the square root of the sample size [6, pp. 195–216]. The standard deviation of the

Vol. 6
1978

population, in this case the target variable, is the same whether the sample of size n is drawn by means of a single-tiered model (such as a direct probability model) or a many-tiered model (such as would be involved in an ambitious simulation) of the same target variable. For personalist, or Bayesian, inference, an analogous argument can be made in terms of likelihood functions and priors. The general argument does not depend on the parent distribution's being normally distributed.

Thus, increasing the complexity of a model of a target variable does not, of itself, affect the distribution of the target variable that it is attempting to represent and, therefore, has no effect on the *sampling* error of estimates of the population's parameters. This argument applies equally well to conventional Monte Carlo simulation, though it is by no means universally acknowledged by practitioners of that art, and we have not seen such an argument presented in the technical literature.

Although greater model complexity does not affect the *sampling* error with a given number of simulation trials, the *modeling* error inherent in a model of the target variable may be *reduced* by a more complex disaggregated model. This reduction in modeling error can arise because an effective disaggregation of a problem requires assessments on sub-problems that are much more easily and accurately addressed than the total aggregated problem. (This is a primary motivation for using an extensive tree analysis or Step-Through rather than a simple model.) Thus, increasing the number of steps in Step-Through can frequently offer a reduction in modeling error without affecting sampling error.

3. TESTING THE USEFULNESS OF STEP-THROUGH

Since Step-Through is a new technological idea, it is probably most fruitfully tested by workshop trials, that is, informal reflections based upon trial applications of the methodology in an opportunistic sample of decision situations. This method of testing is not as complex or as rigorous as a controlled experiment and consequently does not produce results that are as definitive. But, at this early stage of development, useful insights can be obtained most economically through workshop trials. If the results of this testing activity indicate that Step-Through is a viable procedure, more controlled experimentation may be indicated for a later time.[2]

Five workshop trials were conducted using a variety of decision problems and decision makers. The problems included military decisions and personal career decisions. Subjects included trained decision analysts and military officers both with and without prior exposure to decision analysis techniques. The trials were carried out after various amounts of pre-modeling effort (structuring the sequence of acts and events, assigning values, and assessing probabilities). At one extreme, Step-Through was used as a first formal consideration of the problem, without any pre-modeling. At the other extreme, Step-Through was used to evaluate and refine a decision model that had already been substantially, although not completely, modeled. This substantial pre-modeling included an elicitation of the complete schematic sequence of acts and events, an elicitation of an influence diagram describing the conditional nature of the probabilities, and an assignment of outcome values. At an intermediate level of pre-modeling, Step-Through was used after eliciting a modal path through the tree for each initial option, identifying the value criteria, and assigning values to outcomes.

The pilot testing results supported our initial assumption that Step-Through can offer a considerable economy of elicitation and computational effort over a complex extensive tree. The amount of advantage offered by Step-Through, however, was found to depend critically upon the proper implementation of the method, especially with respect to the proper amount of pre-modeling. That is, both too little and too much pre-modeling of a decision problem was found to affect adversely the advantage of Step-Through. When used with little or no pre-modeling, the Step-Through procedure produced models that were grossly inaccurate simplifications of the decision situation. On the other hand, a high degree of pre-modeling produced structurally sound models, but the pre-modeling itself required a large elicitation

[2] The interested reader can find a more complete discussion of the range of possible evaluation techniques and the reasons for choosing each in [5].

effort and thus reduced the advantage of Step-Through. When a moderate amount of pre-modeling was performed, however, a structurally sound model was obtained at a modest elicitation cost. Overall, when properly implemented, Step-Through was found to reduce the total elicitation and computational burden of an extensive tree analysis of comparable complexity by over 50%. We feel that this advantage can be expected for most moderately to highly complex decision models (models containing ten or more tiers of action or event nodes).

The workshop trials also proved to be very useful in answering four implementation questions:

1. How much pre-modeling should precede the Step-Through procedure?

2. Should other forms of decision analysis be attempted before applying Step-Through?

3. How much does the decision maker need to be involved in a Step-Through modeling effort?

4. How much does a decision analyst need to be involved in a Step-Through modeling effort?

Firstly, the tests indicated that the Step-Through procedure is most effective when it is used after a moderate amount of pre-modeling. The pre-modeling should consist of an elicitation of a modal or most likely path through the tree for each immediate action alternative, an elicitation of the criteria of value in the decision situation, and an assignment of values to the events that directly contribute to the overall valuation. Further, our tests indicate that moderate pre-modeling helps to reduce the amount of confusion that the decision maker experiences during the Step-Through procedure, and this enables him to provide inputs that accurately reflect his perceptions of the decision situation.

Secondly, the tests indicated that a full-scale Step-Through should not be used as the first formal method of analyzing a decision problem. Step-Through was found to require a large commitment of resources in terms of monetary expense, a decision analyst's time, and the decision maker's time. Consequently, a full-scale Step-Through analysis can be justified only in decision situations that cannot be adequately addressed by less costly analytic techniques, such as a direct probability model[3] or Modal Step-Through.

Thirdly, the workshop trials indicated that the decision maker needs to be heavily involved in the process of building and using a Step-Through model. His heavy involvement increases the chances that an accurate model will result and that he will understand the output of the model. With this involvement, Step-Through effectively highlights the key considerations of a decision problem in a way that facilitates their precise modeling and estimation. This results in an accurate model of the decision problem. Heavy involvement is also necessary to communicate the results of the model, that is, the probable consequences of selecting each initial option. In addition, the tests indicate that heavy involvement of the decision maker in Step-Through analysis serves an immediate training function by increasing the decision maker's understanding of the decision situation. The trials also indicated that a decision analyst needs to be heavily involved in the Step-Through modeling procedure to guide the decision maker through the sample paths and to provide an accurate interpretation and explanation of the output.

Finally, the workshop trials indicated that the prototype computer implementation of Step-Through [7] requires a substantial amount of development, especially in the area of a streamlined implementation, before it is suitable for a production environment.

4. CONCLUSIONS

Our limited development, application, and testing allow us to make only tentative conclusions about the usefulness of the Step-Through variant of Monte Carlo simulation for structuring a decision model, pruning a decision tree, and evaluating initial action options. Because Step-Through offers elicitation economies over the extensive tree variant, and because it retains the disaggregation properties of the

[3] For a discussion of how alternative decision analytic approaches compare on cost, precision and other measures of performance, see [4].

extensive tree, Step-Through is a very effective procedure for developing an accurate structure of a complex decision situation. Because the Step-Through procedure effectively identifies the key parameters in a decision model, it is a useful procedure for pruning an extensive decision tree by eliminating superfluous branches. This, together with Step-Through's computational economy, implies that it is a useful technique to evaluate initial options as a basis for choice among them.

The application and testing activity also revealed that Step-Through is a useful procedure for training decision makers for contingent situations. In particular, its ease of operation and limited demands, compared with conventional Monte Carlo simulation, enable a decision maker to simulate easily an uncertain future situation and to develop an understanding of the situation and a 'feel' for the best response to it.

On the other hand, Step-Through does not appear to be a promising variant to use as an initial formal approach for modeling a decision situation, for evaluating subsequent acts in complete decision strategies, or for use without expert supervision. The workshop trials have shown that Step-Through requires a large commitment of resources and thus is justified only after simpler modeling techniques prove to be inadequate. In addition, Step-Through appears to produce over-simplified structures of complex decision models when no pre-modeling is performed. Since only a small fraction of the sample paths through the implicit decision tree will contain the same setting for subsequent acts, a Step-Through analysis to evaluate complete strategies may require a sample that is too large to be practical. Since the Step-Through procedure is fairly complicated and its output is difficult to interpret, a good Step-Through analysis requires a high degree of participation from both the decision maker and a trained decision analyst.

Finally, the testing that has been completed suggests that the viability of the Step-Through variant warrants further testing, development, and application, both in controlled experimental settings and in applications to real decision problems. Readers who are interested in carrying out such tests, development, or applications should request a copy of [7] which contains a much more detailed description of the Step-Through simulation technique illustrated in the context of an actual decision problem.

REFERENCES

1. Brown RV (1975) *Modeling Subsequent Acts for Decision Analysis.* Technical Report 75–1. Decisions and Designs, Inc., McLean, VA 22101, USA.
2. Brown RV, Hoblitzell CM, Peterson CR & Ulvila JW (1974) *Decision Analysis as an Element in an Operational Decision Aiding System.* Technical Report 74–2, Decisions and Designs Inc.
3. Brown RV, Kahr AS & Peterson CR (1974) *Decision Analysis for the Manager.* Holt, Rinehart & Winston, New York.
4. Brown RV & Ulvila JW (1976) *Selecting Analytic Approaches for Decision Situations: A Matching of Taxomomies.* Technical Report 76–10, Decisions and Designs Inc.
5. Brown RV & Watson SR (1977) *Testing Procedures in the Designs of Management Systems: Some Methodological Reflections.* Technical Report 77–1, Decisions and Designs Inc.
6. Larson HJ (1969) *Introduction to Probability Theory and Statistical Inference.* John Wiley, New York.
7. Ulvila JW, Brown RV & Randall LS (1976) *Step-Through Simulation: A Method of Implementing Decision Analysis.* Decisions and Designs Inc. Copies of this report are available through the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161, USA, Tel. (703) 557–4650. Persons requesting copies should specify its Government Accession Number: AD A036969.

ADDRESS FOR CORRESPONDENCE: *Jacob W Ulvila Esq.* 1622 *Worcester Road, Apartment* 209-*B, Framingham, Massachusetts* 01701, *USA.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 33–42, 1978. Pergamon Press. Printed in Great Britain

The Benefits of Planning with Normative and Heuristic Approaches[1]

PAUL C NUTT

The Ohio State University

(Received March 1977; in revised form June 1977)

The effectiveness of planning groups that used normative planning approaches and heuristics were investigated. The quality and innovation in plans developed using each planning approach were compared. Two studies were carried out in three organizations. In the first study, a 'behavioral' planning approach and a heuristic were used in a single organization. In the second study, a 'systems' planning approach and the heuristic were applied in two organizations. The field experiments demonstrated that the normative planning methods produced plans with superior quality and higher information levels. No differences in plan innovation were observed. Some guides for practice were derived from these findings.

1. RESEARCH ISSUES

THE SUPERIORITY of Normative planning approaches over Heuristics (unstructured discussions among experts) is a common theme in the management science literature. Normative methods are thought to provide a mechanism which stimulates experts in a planning group to provide quality and innovative ideas [3]. For example, specifying problems [5] or objectives [12], prior to considering solutions and evaluating their merits, is thought to be superior to merely listing solutions [12]. Others contend that these mechanisms are misleading and ineffective and advance lengthy justifications for heuristics [7, 9].

Empirical evidence which supports the use of either Normative or Heuristic Methods is sparse. The typical study has manipulated contextual factors. The power of the planner, the perceived impact of the plan on the agenda of decision makers, the involvement of users in planning, or the time allocated to planning have been studied—not methodology [6, 14]. To contrast these methods of planning, the plans made using Heuristics must be compared, using several criteria, with the plans drawn by Normative approaches. This paper reports research which compared Heuristics and Normative approaches used to guide the planning committees of human service organizations.

2. NORMATIVE PLANNING APPROACHES AND THEIR RATIONALE

Ramo [16] observes that health, education, environment, and transportation systems often have bloated budgets and poor performance records. Ramo contends that these problems stem from the Heuristics used by the planning groups to devise these services. He associates the high performance of telecommunications, space, and military weapons systems with the Normative methods (systems approaches) used to design these systems. Churchman [2] chides the 'anti-planning' attitude which often accompanies planning groups charged with public sector planning. Again, Heuristics are cited as a likely cause of poor plans. Churchman outlines several approaches to planning, including those of behavioralists and economists, and points out the advantages of the normative elements in each approach [2].

[1] This research was supported by a grant from the National Center for Health Services Research (RO 1413) and the Ohio State University.

Springer [18] cites the advantages of 'strategic planning', a process which *requires* key organizational units to formulate plans annually. The plan must contain specific features, such as product demand forecasts and manufacturing contingency plans for each level of demand. These requirements were found to stimulate use of normative planning procedures. Variegated plan requirements appear to render heuristics ineffective and cause the adoption of normative planning methods.

Thompson [19] suggests that planning groups in organizations tend to develop plans that are very similar to the current plan. Departure from current practices seldom occur, even when current practices were known to have problems. Cyert and March [4] found that 'conspicuous' solutions tend to be adopted by planning groups, regardless of their quality. In each of their widely cited case studies the planning approach was essentially Heuristic. The frustration which accompanies trial and error planning (Heuristics) may cause the adoption of 'fatigue solutions' which, in turn, may explain the low level of planning innovation cited by Thompson. Simon and Newell [17] found that people tend to use a 'progressively deepening' approach (a personal Heuristic) when planning. When applying Heuristics, people became prematurely committed because of past actions and choices. Heuristics were found to be less effective than a 'scan and search' (Normative) strategy, that explores a wide array of options. Thus, Heuristic approaches are often used by planning group members, and they seem to be ineffective.

An increase in the number and diversity of ideas occurs when planning groups use Normative planning methods [1, 11]. Structured agendae and formal planning procedures appear to enhance the likelihood that good ideas become visible to a planning group [20]. For example, silently listing ideas prior to discussion has been found to increase 'search behavior' which stimulates quality and innovative plans [20]. In other studies, cited by Maier [10], the key ingredient in successful planning meetings appeared to be a structure: systematically gathering ideas, allocating different parts of the meeting to ideation and evaluation, increasing the time available for planning, and following a set strategem during the meeting. Nutt [13] found that Heuristics used to develop health service plans, with service users as planning group members, made plans lacking in quality and innovativeness. When experts served as planning group members, the Heuristically derived plans dramatically improved. Because most planning groups are made up of experts, mixing the type of planning group participants and methods rendered these findings inconclusive.

3. ADAPTIVE PLANNING USING HEURISTIC APPROACHES

Lindbloom [9], Gore [7] and others believe that planning should be Heuristic. Planning groups are formed to permit planners to monitor the sentiments and viewpoints of key community actors and to infer their requests for action. If these requests seem valid and mandate action, the plan sponsor implements them with minimal study. Thus, Heuristic planning does not follow to any set pattern or internally consistent logic. 'Heuristic planning', while fatalistic and highly political, aptly describes the conduct planning groups, in certain settings, such as planning for government sponsored social programs [20].

The advocates of Heuristics also believe that Normative methods have many serious limitations, particularly when applied to planning public sector services. For example, critical specifications, such as the mission and scope of a project, can be difficult to determine. Thus, critics of Normative methods contend that important problems are seldom dealt with in even a remotely comprehensive manner by a planning group because, in part, they are difficult to define. A failure to agree on goals to be met or problems to be solved, renders the search for solutions ambiguous [19]. The investment in data and analysis becomes extensive and the costs of planning are difficult to justify.

Finally, critics contend that Normative planning methods are unable to capture fully or to describe complex social problems with their intertwined sociotechnical components. Hoos [8] points out that systems procedures applied to social problems exclude all but the quantifiable factors. Cultural, value, and political aspects of the problems are typically ignored. The pseudo-scientific language of Normative methods is believed to keep these oversights

obscure. "Muddling through is probably safer... (and) the unstructured participation of experts (in planning) is preferred [8]".

4. RESEARCH QUESTIONS

A field experiment was designed to compare the attributes of heuristic and normatively derived plans, using the methodology of past studies [13, 14]. Answers to the following questions were sought:

1. Do Normative methods produce plans that have more quality and information, and require less meeting time, than plans made by Heuristics?
2. Are Normative or Heuristic methods preferred by planning group members?

In the discussion that follows, the steps and procedures of the Normative and Heuristic planning methods tested in this research are described and the measures of plan quality, acceptance, innovation, and cost are defined. Table 1 compares some attributes of the Normative and Heuristic planning methods.

5. THE NORMATIVE PLANNING METHODS USED IN THE EXPERIMENT

A behavioral approach

In the Behavioral approach plans were based on perceptions of the needs of clients. As a first step, the problems of a client group were identified using 'nominal group' process [5]. The nominal group format requires a formal meeting which begins with group members listing their views of client or user problems without discussion. Next, the problems were transcribed by the group leader, recording the problems from each group member one at a time. 'High priority' problems were ranked by the group. Following a break, the nominal group format was used to identify *solutions* to overcome these *priority problems*. Each plan was made up of the five priority solutions selected by the participants.

A systems approach

To initiate planning in the systems approach, a hierarchy of plan objectives was developed [12]. First, ideas for objectives were solicited. Next, the recorded objectives were reviewed to identify the objective with the smallest scope. An array of objectives was developed by gradually adding objectives of larger scope. A Hierarchy of objectives resulted, which range from trivial to unachievable. A particular objective was selected by the group to guide its planning activities. After a break, the nominal group process was used to identify ways to achieve the selected objective. As before, the plan was made up of the five priority solutions selected by the participants.

TABLE 1. A COMPARISON OF HEURISTIC AND NORMATIVE METHODS

Some key characteristics	Normative approaches	Heuristic approaches
Group processes	Nominal Group management methods	Interacting group
Planning procedure	(a) problem specification (b) define problems or objectives (c) explore solutions (d) formulate plans (e) evaluate (f) prioritize solutions (g) planning group members assess their efforts	(a) problem specification (b) list solutions (c) prioritize (d) planning group members assess their efforts
Key criterion	Plan quality and acceptance	Minimal cost and responsiveness to demands
Rationale	A logic structure	Adaptation of existing ideas
Planning group members	Experts	Experts

6. THE HEURISTIC PLANNING APPROACH USE IN THE EXPERIMENT

To model elements of Heuristic planning behavior, planning groups were presented with a problem statement, identical to that used for the behavioral and systems approach, and asked to recommend solutions. The experimenter merely recorded solution ideas from the group until the group's ideas were exhausted. The plan was made up of the five priority solutions selected by the planning group members.

7. THE RESEARCH SETTING

This research was conducted in three organizations: a state department of health, a university health service, and a health systems agency. The participating organizations selected the planning topics and recruited all participants. The experimental planning groups were made up of the sponsoring organization's staff or their constituents. Evaluation groups were drawn from each organization's staff and decision groups.

Study I: *a single organizational site*

Environmental health planning task. To compare the effect of Normative and Heuristic methods, the 'Behavioral' and the 'Heuristic' methods were used to study ways of improving regulatory activities in the Bureau of Environmental Health, for a State Department of Health. The Bureau's regulatory activities include the inspection of businesses, such as hotels and restaurants, and commodities, such as water supplies, sanitation, radiation levels, and milk, to detect health hazards. Other functions include the prosecution of violators and identification and dissemination of ways to avoid health hazards. The Bureau had a central staff and a district staff, located in one of eight multicounty regions. Both groups had several regulatory roles. The executive staff of the Health Department believed that conflicts between district and central staff had rendered the inspection, prosecution, and consultation functions ineffective.

Forming the planning groups. Experimental planning groups were composed of the staff of the Bureau of Environmental Health in the State Department of Health. A Section Chief, a District Office Administrator, and inspectors, including Sanitarians, Environmental Engineers and Plumbers, were represented on each planning group (Table 2). The participants were recruited by the Assistant State Health Officer to advise the Executive Staff on ways to improve environmental health regulation. Participants were told that priority ideas would be considered in any reorganization plan. The research design for Study I is shown on Table 2.

Study II: *the impact of Heuristic and Normative planning methods in several organizations*

Planning was conducted in several organizations to consider the effects of Normative and Heuristic planning methods in several settings. A university health service (UHS), and a health systems agency (HSA), in addition to the state Health Department in Study I, participated. Each applied the Systems and the Heuristic methods. The health systems agency selected two planning topics: the adequacy of primary health care arrangements in an urban area and a county-wide home health care program. The university health service requested a study of the needs for student dental services.

Dental care planning task. Dental care became an issue for the university health service (UHS), when the university administration tried to get the UHS to qualify for 'Health Maintenance Organization' funds. To obtain these funds, dental services must be added to the UHS benefit package. UHS services were prepaid in student tuition so these services could be provided by increasing tuition. The UHS chief executive hoped to learn what types of student dental services could be justified.

TABLE 2. STUDY I: RESEARCH DESIGN

Planning groups*	Method
9, 10, 11	Normative (behavioral)
12, 13, 14	Heuristic
Planning group membership* (groups: 9, 10, 11, 12, 13, 14)	
District Office Administrator Officer Section Chief Sanitarian Plumbing Inspector Environmental Engineer	

* Each planning group had five members and was used to formulate a single plan for the experiment.

Vol. 6
1978

TABLE 3. STUDY II: RESEARCH DESIGN

Planning group	Experimental trials: Research site	Method
1, 2, 3, 4	Health Systems Agency	Normative (Systems)
7, 8	University Health Service	Normative (Systems)
12, 13, 14	State Department of Health	Heuristic
5, 6	Health Systems Agency	Heuristic

Planning group membership*

Primary Care—(1, 2, 6)
- Pediatrician
- Family Practice Physician
- Emergency Room Nurse
- Hospital Administrator
- Public Health Nurse

Home Care—(3, 4, 5)
- Physical Therapist
- Public Health Nurses (3)
- Hospital Home Care Coordinator

Dental Care—(7, 8)
- Physicians (2)
- Registered Nurses (2)
- Receptionist

Environmental Health—(12, 13, 14)
- District Administrative Officer
- Section Chief
- Sanitarian
- Plumbing Inspector
- Environmental Engineer

* Each planning group had five members and was used to formulate a single plan for the experiment.

Forming the planning groups. Each dental care planning group was composed of fulltime staff of the University Health service with several years of experience. Two Physicians, two Registered Nurses, and a receptionist, who assigned patients to physicians or nurses using the patient's descriptions of their health problem, were represented on each planning group (groups 7, 8, in Table 3). The planning group members were recruited by the UHS director who asked them to help him prepare a response to the University Administration's request.

Primary care planning task. Expansion of primary health care services, in a community within the jurisdiction of the HSA, had been demanded by city residents for several years. As a measure of the intensity of these views, plans for a neighborhood health center were under development by several citizen planning groups. These groups had been sanctioned by the Mayor and other Community leaders. Local physicians and hospital administrators were wary of expanding health facilities because several studies had been conducted which found that local health facilities had excess capacity. The HSA would be asked to approve any service expansion and hoped to determine what type of expanded primary care service could be justified.

Forming the planning groups. Participants for the primary care groups were suggested by the leaders of local health trade associations and elected leaders in professional societies. The president of the County Pediatrics Society identified several pediatricians and the president of the State Medical Society recommended several family practice physicians. The associate administrator of each of the three community hospitals participated and recruited some of their emergency room nurses. Public health nurses who served in the area being studied were also recruited. Each planning group was composed of a Pediatrician, a Family Practice Physician, a high-ranking Hospital Administrator in one of the community hospitals, an Emergency Room Nurse, and a Nurse with Public Health background (planning groups 1, 2, 6 in Table 3). Each planning group was asked to suggest economical ways to expand Primary Care Services.

Home care planning task. The Visiting Nurse Service, charity funded, and the County Health Department, publically funded, provided Home Health Care Services in a county under the jurisdiction of the HSA. The controlling boards of these organizations, elected county officials and the United Way Board, were considering how they could improve the coordination of these Home Care Services in the county. The HSA was asked, by the agency sponsors, to study how Home Care Services were provided and to make recommendations to improve the delivery of these services.

Forming the planning groups. Members of the Home Care planning group were selected from the staff of the agencies providing home care services and the home care coordinator in each hospital that served the same area. Each Home Care agency director recruited their staff to serve on the planning groups. The local hospital administrators recruited their home care coordinator. Each of the three Home Care planning groups were composed of one Physical Therapist, and Administrator, and three Public Health Registered Nurses. (Planning groups 3, 4, 5 in Table 3.)

8. CONTROLS USED IN FIELD SETTINGS

The planning meetings were initiated by a staff member of the sponsoring organization, who described the planning topic and the purpose of the meeting. This provided planning groups with comparable sets of information and similar expectations.

The experimentor was introduced as a consultant, designated by the organization's chief executive to act as the planning group's chairman. This introduced certain experimental risks, but is preferable to having the group leader elected by each planning group or designated by the executive. Using a single leader reduces the chance that skill and other leader characteristics could influence results. It also allowed the experimentor the latitude to carefully follow the protocol required by each planning method. This protocol prohibited substantive comments by the leader.

The research designs (Tables 3 and 4) describe the experimental contrasts, replications, and how each planning group was used in each study. Planning groups with similar expertise and skill profiles were formed to permit replication. A process of random assignment was used to make the experimental planning groups as much alike as possible. Participants in these groups were likely to have neutral views of the method applied. Each planning group was used *once* in the experiment to provide one independent observation of the application of a particular planning method to a particular planning task.

9. MEASURES USED TO EVALUATE THE PLANS DRAWN BY THE EXPERIMENTAL PLANNING GROUP

The plans made by each planning group were evaluated using measures of their 'quality' and 'information', the time allocated to planning, and the reactions of members of the planning groups. Quality was determined by 'experts' and a decision making body of the participating organizations. Three expert

TABLE 4. COMPARING THE EFFECTIVENESS OF HEURISTIC AND NORMATIVE PLANNING METHODS

	Plan quality*				Plan information**				
	Agency decision group	Experts			Agency decision group	Experts			
Study variables (planning methods)	Adoption	Technical Features	Acceptance Attributes	Adoption	New Information	New Information	Elaboration	Number of solutions	Meeting time (hr)
Study I									
Normative	3.2	3.0	3.1	3.5	5	7	21.3	30.7	2.8
Heuristic	2.5	1.3	2.6	2.2	5	4	12.7	16.3	1.7
Significance	0.10	0.05	—	0.05	—	—	0.10	0.01	0.10
Study II									
Normative	3.2	3.3	3.2	3.0	4.8	1.7	19.5	23.5	2.2
Heuristic	2.9	2.1	2.7	2.4	4.6	3.0	18.0	11.2	1.6
Significance	—	0.05	0.10	0.10	—	—	—	0.10	0.005

* Scale
0 = No Value;
1 = Marginal;
2 = Low Priority;
3 = Important (Threshold);
4 = Very Important;
5 = Essential.
An average rating of 3.0 was required to take action.

** Scale: frequency counts.

evaluators were selected for each planning topic. Dentists, who had been accepting students needing emergency dental care on a referral basis from the University Health Service for more than 3 years, evaluated the dental care plans. The environmental health expert evaluators included the current and former director of research for the Health Department and a university based expert in public health. Nursing School faculty members evaluated the Home Care plans. University medical school faculty, specialising in Primary Care medicine evaluated the Primary Care plans. A decision making body from each of the participating organizations evaluated the plans of the experimental planning groups. The University Health Services' executive director, associate director, director of nursing, and controller made up the agency decision group for 'dental care'. The environment health 'agency decision group' was composed of the state health officer, the assistant state health officer, and the director of budget. In both instances, the 'agency decision group' had served as the agency's 'executive committee' and often made decisions similar to those required by the research project. The Health Systems Agency selected their 'County Health Resources Committee' (Primary Care), and their '*Ad hoc* Home Health Care Committee' (Home Care), to determine which plans merited adoption.

An evaluation procedure was devised to model group decision making behavior. The five priority plan components, from each planning group, were randomly listed without reference to their source or their priority. The evaluation groups were asked to rate the plans. They were told their ratings would be used by the plan's sponsor in selecting a course of action.

Three criteria were used to evaluate plans: (1) technical features, the extent to which plans used optimum technology and capabilities and dealt with priority needs; (2) acceptance attributes, the level of social and political factors that permit or limit implementation of the plan; (3) adoption, those plans the agency should try to implement. The following scale was used for each criterion:

0 = No Value (no good features);
1 = Marginal (a few good features);
2 = Low Priority (several good features, but insufficient priority to adopt);
3 = Important (adopt, moderate priority);
4 = Quite Important (adopt, high priority);
5 = Essential (adopt, very high priority).

This rating scale was behaviorally anchored to induce interval scale properties in the evaluation data. The zero point and scale increments were carefully defined, so the evaluators would associate these intervals with possible sponsor's action. The evaluators were told that an average rank of '3' suggests that the plan component merits presentation to the plan sponsor.

The evaluation groups used the scale and assigned a number to plan components. The average rankings for each plan component were computed and displayed for group reference. The evaluators were reminded that an average rating of '3' was required for the plan component to be presented to the plan sponsor. After this discussion the ranking process was repeated. These final rankings were averaged to provide three measures of quality: the plan's technical features, acceptance attributes, and adoption potential.

A measure of innovation was obtained by having the evaluators designate each plan component that provided 'new' information, when the plans were first reviewed. *New Information* was defined as 'a situation, need, or idea not previously recognized or tried'. The number of times plan components, from a particular planning group, were designated 'new' by experts and the agency decision groups provided two indices of innovation. Another indicator of information was constructed by counting the number of times plan components were elaborated. *Elaboration* was defined as identifying a target group; specifying a desired action, like a pilot program or an action plan; identifying the results sought by the plan, such as efficiency or information; contrasting current affairs and ideal states; defining obscure terms; or specifying priorities among solution elements. Finally, the total *number* of plan components suggested by each planning group was tabulated and the *length* of each meeting was recorded.

Normative methods may have limited application if planning group members perceive the results from Normative methods to be poor or if the method was thought to be unpleasant. A questionnaire was used to identify the participants' satisfaction with the Heuristic and Normative planning meetings, their evalu-

ations of the planning group leader (experimentor), their perceptions of value of their plan, and whether they would advocate using either the Heuristic or the Normative method in future planning meetings. The questionnaire used a behaviorally anchored rating scale (see Table 5) so that responses would have interval scale properties.

10. FINDINGS

The results of the experiment were analyzed with standard one way analysis of variance procedures. The measures of quality, information, participant reactions, and lengths of meetings were the dependent variables in this analysis. Applying either the Heuristic or the Normative planning method was the independent variable.

As shown on Table 5, 'quality' measures for Study I suggest that the Normative planning method was superior to the Heuristic method. Similarly, in Study II the Normative method consistently outperformed the Heuristic method. In both studies, the differences were statistically significant. (Statistical interactions between the Heuristic and Normative methods and the planning topic were not observed.)

The magnitude and the degree of agreement across several independent indicators of plan effectiveness give these findings special significance. This stems from the instructions given to evaluators. Evaluators were told that *only* plans with an average rating of 3.00 or more would be reviewed by the plan sponsor. Thus if plans were rated 3.00 or less, evaluators would perceive them as having little value for the sponsor.

The quality indices for the *Heuristically* derived plans ranged from a low of 1.3 to a high of 2.6, in Study I, and from 2.1 to 2.7, in Study II. In contrast, *Normative* methods produced plans with ratings ranging from 3.0 to 3.5 in Study I, and 3.0 to 3.3, in Study II. The *Magnitude* of these differences in plan quality suggests that differences have operational, as well as statistical significance. Moreover, Normative methods were found to provide superior results when applied in three separate organizations for four planning tasks using distinct planning groups each with a unique membership profile. Thus, the benefits of Normative methods seem robust because these benefits were demonstrated in several environments.

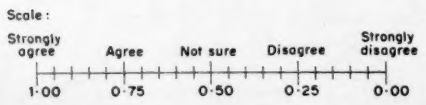
The 'information' measures, also shown in Table 4, suggest that Normative methods when used for planning, provide a larger number of ideas and each of these ideas tended to be more elaborately described. Normative methods were found to have these attributes in *both* Study I and Study II. In contrast, the proxy measures for innovation, new information, found little difference in the number of novel ideas produced by the Normative and Heuristic planning methods. The information measures suggest that Normative methods provide a more detailed description of the plans or actions recommended by planning groups and provide more than double the number of distinct ideas for a group to consider during planning. Normative methods did *not* seem to enhance the novelty of these ideas.

Vol. 6
1978

TABLE 5. REACTIONS OF PLANNING GROUP MEMBERS TO THE HEURISTIC AND NORMATIVE METHODS

	Study I		Study II		
	Normative method	Heuristic method	Normative method	Heuristic method	Statistical significance
Role satisfactions*	0.69	0.74	0.74	0.72	—
Leader evaluation*	0.84	0.88	0.79	0.83	
Perceptions of solution quality*	0.79	0.81	0.77	0.77	—
Use method again*	0.73	0.73	0.67	0.64	—
Increased confidence in organization	0.63	0.65	0.59	0.61	—
Estimates of plan success	0.78	0.74	0.59	0.68	0.10(II)
Mean*	0.75	0.78	0.73	0.73	—

* Composite Measures, formed by averaging several responses to questions in post meeting survey.



As shown in Table 4, the Normative method required 40% more time to apply in Study I and 25% more time to apply in Study II. Planning activities that use Normative methods will tend to cost more than Heuristic approaches because longer meetings are required. But a substantial increase in plan quality in terms of its sophistication and acceptance, an increase in the number of topics considered, and an increase in the plan detail can be expected when using normative methods. The additional time used by normative methods lead to an increase in plan information levels. This, in turn, may have aided normative methods in producing higher quality plans. Thus, the benefits associated with Normative approaches seem to justify the cost.

Table 5 provides a summary of the reactions of planning group members to their planning session. In both Study I, and Study II, the Heuristic approach tended to be preferred. Only differences in the rating of the 'potential for plan success', in Study II, reached statistical significance. The remaining measures were not statistically significant. The satisfactions of members with their planning role, their plan's quality, and the prospect of using each of the methods again were *not* adversely effected when normative methods were used. The applications of normative methods in several organizations did not produce negative reactions in the planning group members.

11. IMPROVING THE EFFECTIVENESS OF PLANNING GROUPS

Many management scientists advocate the universal adoption of Normative planning approaches. According to this view all planning tasks, ranging from problem definition to implementation, can be improved if normative approaches are used. Another portion of the Management Science community believes that normative approaches are ineffective and their costs can not be justified. This research was designed to aid the administrator and the MS practitioner to sort through these claims and counter claims. The finding of the study may have special significance for both practitioners and researchers with an interest in the practice of planning:

Normative planning methods, applied to the planning of human services, were found to increase plan quality. This confirms the findings of several retrospective studies. In these studies, organizations that used 'formal' planning (eg., using specific procedures), were superior to organizations using 'informal' planning, using indicators such as earnings per share as criteria. (6) The findings of this study, coupled with the retrospective studies have special significance because Heuristic methods are widely advocated and because trial-and-error represents the behavior of many planning committees and groups. As a result, opportunities for change and the good will of participants is lost in unproductive planning activities. Merely using a simple technique, like a nominal group, improves productivity and the participants' sense of accomplishment. Finally, plan quality seems to improve when planning topics are phased; considering problem formulation, solution generation, and evaluation sequentially.

A key issue in planning is innovation: new ideas that offer opportunities to improve the operations and the products or services of an organization. As a result, the innovation levels in plans prepared by normative and heuristic planning methods have considerable interest. This research demonstrated that both the number of ideas and the description of important contingencies increased when Normative planning approaches were used. In particular, Normative approaches produced more ideas and described these ideas in more detail by specifying target groups, results sought, ideal states, priorities and the like. But differences in the number of novel ideas were not observed. Applying Normative methods to planning tasks seems to increase information levels, but not innovation.

Administrators in the field have been reluctant to use normative planning methods because they channel and restrict discussion (4). As one administrator remarked, "physicians will never accept these (normative) methods of planning." Normative methods are believed to structure planning to an extent unacceptable to many planning group members. These fears seem unwarranted. This experiment suggests that planning with normative methods will not produce negative participant reactions. The members of the experimental planning groups, exposed to normative and heuristic

methods, were equally satisfied with their role, had nearly identical perceptions of the quality of results produced, and indicated that they were as likely to use one method as the other in future planning activities.

REFERENCES

1. Abernathy WS & Rosenbloom RS (1969) Parallel strategies in developmental projects. *Mgmt Sci.* **14**(10), B-486, B-505.
2. Churchman CW (1968) *The Systems Approach.* Dell, New York.
3. Churchman CW (1971) *The Design of Inquiring Systems.* Basic Books, New York.
4. Cyert RM & March JG (1963) *A Behavioral Theory of the Firm.* Chap. 4, Prentice-Hall, Englewood Cliffs.
5. Delbecq A & Van de Ven A (1971) A group process model for problem identification an program planning. *J. Appl. Behavl Sci.* **7**, 466–492.
6. Filley AC & House RJ (1969) *The Management Process and Organization Behavior.* Scott Forsman, Glenview, IL.
7. Gore WJ (1964) *Administrative Decision Making: A Heuristic Model.* John Wiley, New York.
8. Hoos IR (1972) *Systems Analysis in Public Policy: A Critique.* University of California Press, Berkeley.
9. Lindbloom CE (1965) *The Intelligence of Democracy: Decision Process Through Adjustment.* Free Press, New York.
10. Maier NRF (1970) *Problem Solving and Creativity: In Individuals and Groups,* Brooks-Cole, Monterey, CA.
11. Mason RO (1969) A dialectical approach to strategic planning. *Mgmt Sci.* **18**(8), 475–487.
12. Nadler G (1970) *Work Systems Design.* Irwin, New York.
13. Nutt PC (1977) An experimental comparison of the effectiveness of three planning methods. *Mgmt Sci.* **23**(5), 499–511.
14. Nutt PC (1976) The merits of using consumers and experts in planning: a Field experiment. *Acad. Mgmt Jl* **19**(3), 378–394.
15. Nutt PC (1976) Models for decision making in organizations and some contextual variables with stipulate optimal use. *Acad. Mgmt Rev.* **2**, 84–98.
16. Ramo S (1969) *Cure for Chaos.* McKay, New York.
17. Simon HAA & Newell A (1971) Human problem solving—state of the art in 1970. *Am. Psychol.* **26**(2), 145–159.
18. Springer CH (1973) Strategic management in General Electric. *Op. Res.* **21**(6), 1177–1199.
19. Thompson JD (1967) *Organizations in Action.* McGraw-Hill, New York.
20. Van de Ven A (1972) An applied experimental test of the nominal, delphi, and interacting decision making process. Ph.D. Dissertation, University of Wisconsin, School of Business, Madison.

ADDRESS FOR CORRESPONDENCE: *Dr Paul C Nutt, The Ohio State University, the Graduate Program in Hospital and Health Services Administration,* 1583, *Perry Street, Columbus, Ohio* 43210, *USA.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 43–51, 1978. Pergamon Press. Printed in Great Britain

Management Science in the United States and Europe: A Decade of Change in the Literature

AJ MICHEL

Boston University

SE PERMUT

Yale University

(Received August 1977)

This paper focuses upon the investigation of change in the management science literature from 1965 to 1974. Both subject areas and reported areas of implementation are investigated. The study classifies the content of each article published during the past decade in the following set of European and American journals: United States—*Management Science*; British—*Operational Research Quarterly; and German—*Zeitscrift für Operations Research*. A comparison of the results along both inter and intra country dimensions is presented.**

Vol. 6
1978

INTRODUCTION

THE FOCUS of this paper is the investigation of change in the management science literature during the period 1965–1974. The literature provides the researcher and practitioner with the foundations of theoretical development and the results of previous implementations. Through a longitudinal study of changes in the focus of the literature one can investigate both developing paths of research and successful new areas of implementation.

Since management science was initially developed and utilized in the late 1940's to solve a set of important military problems, both its range and scope have been significantly extended. The analysis of its direction and development are important in assessing the 'state of the science'. However, our knowledge of the progress and resulting changes in the discipline appears to be incomplete. Few studies exist describing either the changing nature of management science research or implementation. There has been a substantial body of knowledge conveyed through the management science literature. Yet our knowledge of the changing development of the subject matter, the changing types of implementation, and the resulting decision payoffs is relatively vague. Urban [6] indicates that only 15% of the articles in *Management Science*: Application "were applied even once in a real decision situation". As he points out, it is difficult to determine the exact level of implementation through the literature because some work is obviously proprietary and other implementations which fail tend not to be reported. A further bias exists in viewing the research reported in the literature as representative of management science research. It must be recalled that the published literature reflects the biases of the editors and to that extent may reflect the normative perceptions of the editors rather than a descriptive representation of the research and implementations in practice. Yet it is the literature which provides the researcher with a primary source of awareness of his research environment and the practitioner with an understanding of the state of applications oriented information and of alternative models previously implemented in an organizational context.

Radnor, Rubenstein and Tansik [3] report a set of strategies to bring about successful implementation. However, relatively little of the

* Now *Journal of the Operational Research Society*.

vast effort devoted to management science research seems to have culminated in implementation and improved organizational decision making. A number of reasons can be postulated why there has been relatively little impact of the management science literature on its practice:

1. The literature is not reflective of organizational needs.
2. Organizations distrust the use of management science models.
3. The assumptions made in models described in the literature are not reflective of the actual environment.
4. The models described in the literature are too complex to be understood and implemented within an organization.

In order to learn more about the literature available to the practitioner (and researcher alike) and to determine if and how the literature has changed over the past decade, a simple accounting of the subject matter is presented. The paper attempts to categorize longitudinally by subject an important segment of the management science literature.

METHODOLOGY

A prestigious management science journal from each of the United States, Great Britain and Germany was selected. Those chosen were *Management Science* (*M.S.*), *Operational Research Quarterly* (*O.R.Q.*), and *Zeitschrift für Operations Research* (Z.O.R.)[1] Each article in these journals between January 1965 and December 1974 was then classified by subject matter and year. This did not, however, include short notes, comments or rejoinders. The subject matter classifications were developed after reviewing each of the articles. In other words there was no preselected set of subjects into which the articles were classified. After reviewing each article, either a new subject category was created or a previous category was assigned.[2] This categorization assessed the subject matter focus of each article, not the methodology or mathematical technique used. To determine if there has been any significant change in the focus of the literature during the previous decade, each classification was divided into two 5-year data sets, 1965–1969 and 1970–1974. A *t*-test was then used to compare the number of articles in each subject category during each 5-year period.[3] The results using a one-tailed test are tabulated for those differences significant at the 0.05 and 0.01 levels.

The number of articles describing implementations in the three reviewed journals is investigated. A tabulation is performed to determine if Urban's reported 15% implementation rate [6] is representative of only the cross-section he utilized or whether the number of reported implementations has been stable over time.[4] Tabulations are developed for articles which:

1. Describe an implementation in a specific environment (hospital, firm, etc.)
2. Do not describe an implementation, but test the reported model using actual data.

Both the actual number of articles reporting an implementation and the percentage of the total articles in each of the three journals is presented.

RESULTS

Table 1 illustrates the number of articles in each of 88 classifications reported in *Management Science* from 1965 to 1974.

The classifications in which the number of articles has significantly increased are: control, functional interfacing (e.g. production–finance), health care planning, manpower planning, national and state planning, networks, planning horizons, reliability, set theory, transportation planning and voting. Those showing a significant number of declines, on the other

[1] Previously *Unternehmungsforschung*.

[2] Each article was assigned to one and only one category.

[3] The use of the proportion of articles in each category to the total number of articles published each year was considered as the test statistic; however the actual number of articles was determined to be more relevant to the analysis of subject matter available to the practitioner or researcher.

[4] Urban, as indicated previously, bases the 15% figure on the "Application Series" articles. This is equivalent to 7.5% on our total article basis assuming the number of Applications and Theory Series articles are approximately equal.

Vol. 6
1978

TABLE 1. CLASSIFICATION OF ARTICLES BY SUBJECT IN *Management Science* FROM 1965 TO 1974

Classification	Year '65	'66	'67	'68	'69	'70	'71	'72	'73	'74	Direction of change and significance
Agriculture, conservation	2			1		2			1		
Bidding			2		1				1	1	
Card-playing							1				
Centralization–decentralization	1			1			1		1	1	
Chance-constrained programming		1	2			1	2	1	2	1	
City and urban planning					1	3		5	3		
Cluster analysis			1								
Communications		1	1								
Computers, EDP	2	3		3	2						−,*
Control theory, practice							1	2	2		+,*
Corporate planning	1	1	1	1	1	2	1	1	1	1	
Cost-benefit analysis		1				1	2				
Cutting stock problem							1				
Decision making		2	1	1	3	4	1	3	1	2	
Decision theory	1	1	1	1			2	1	2	2	
Design			1					2	1		
Diet, menu planning										2	
Dynamic programming	1	1			1				2	2	
Earthquakes										1	
Education	1		2	1	2	13	2	3	4		
Energy planning		3	2	1	1	1	2	2	2	1	
Environmental planning			1				1		8	2	
Family planning								1	1		
Finance	6	13	13	10	12	5	12	9	4	7	−,*
Fire prevention									2		
Forecasting	1	5	2	1	1	1	1		3	2	
Functional characterization		2	1	1		2		3	1		
Functional interfacing (e.g. production–finance)					1	1	1	1	1	1	+,**
Game theory		1	1	3	1	1	1	11	1	2	
Goal programming								1			
Graph theory			3				2	1	1	1	
Growth models				1		1		1			
Health care planning		2			3	1	3	5	7	4	+,*
Human behavior		1	1	1	1						−,**
Human capital										1	
Integer programming	1	5	2	1	2	2	3	4	4	2	
Inventory	5	7	8	4	10	6	7	7	11	8	
Library planning			1						1	1	
Linear programming	7	5		2	2	4	4	2	2	1	
Logistics		4	2	1	1	2		5	3	3	
Managerial research			1			1					
Manpower planning	1	2	1	1	1	1	2	3	2	2	+,*
Marketing	3	5	5	8	3	2	16	4	4	6	
Markov models		1	2	1	2	2	1	3	5	1	
Military planning	1		2								
Min–max theory					2			2		1	
MIS	1		2	2		2	2	1	1	4	
Model building			1					1			
MS discussion, review	6	1	2	1	1	1		3	2		
Multiple criteria problems	1	2			1			1	1		
National, state gov't planning						2	3		2		+,*
Negotiation					2					2	
Networks	1	1		2		3	6	2	4	4	+,**
Newspaper subscriptions				1							
Non-linear programming	2	7	7	3	5	7	4	6	5	4	
Optimization theory		1				1		1			
Organization structure, behavior	2	3	3	1	1	3	4	2	3	1	
Pattern classification								1			
Personnel management			2								
PERT, CPM	1	2	3	1	2		2	1	1		−,*
Planning horizons				1	1	1			2	2	+,*
Plant layout								1			
Price fixing			1								
Probability, statistics		1		1				2	1		

TABLE 1. *Continued*

Classification	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74	Direction of change and significance
	Year										
Production planning	5	3	4	2	4	6	8	2	7	5	
Productivity	1										
Product-mix problem						1					
Queuing	2	1	2	1	4	4	2	4	8	1	
R&D	2		3		3		2	3	5	2	
Ranking			1		2		2			7	
Reliability, replacement, maintenance	3	4	2	1	3	2	5	5	4	5	+,*
Resource allocation		1	1		2	1			1	1	
Scheduling, sequencing, assignment	6	8	4	5	6	9	5	10	14	10	
Set theory						1	1			2	+,*
Simulation	1	2	2	2		1	3	1	1	3	
Space travel			1							1	
Steel production							1		1		
Stochastic programming			2	1				1	2	2	
Systems analysis	1		1	1	3		1			2	
Technology		1	1								−,*
Textile, style goods planning	1	1	1				1	1	2		
Time series		2				1					
Tourism									1		
Transportation planning			1			1	2	3	1		+,*
Traveling salesman problem		1							1		
Utility theory	1	1	1	1			2	1	2	1	
Voting						2			1	1	+,*
Weather			1							1	
Total	71	109	109	72	94	108	126	136	157	120	

⁻ Number of articles in classification from '70 to '74 significantly fewer than from '65 to '69.
⁺ Number of articles in classification from '70 to '74 significantly greater than from '65 to'69.
** Significant at the 0.01 level.
* Significant at the 0.05 level.

hand, are PERT and CPM, computers, finance, human behavior and technology.[5]

What seems surprising about the results are not so much the categories which have increased in frequency, but those having declined. Computers, finance, human behavior and technology are certainly subjects receiving substantial current attention. Yet each of these areas has developed a literature of its own. In fact, the emergence of a distinct literature for each may well be a plausible reason for their decline in attention in *M.S.* The declining interest in PERT and CPM reflected in the literature may be attributable to their early development and abundance of early research.

Looking at the classifications which have not significantly increased in frequency, one finds surprisingly that environmental planning, city and urban planning, and energy planning are among those not having increased. Although the difference between the mean number of articles appearing each year from 1965–1969 and from 1970–1974 is relatively high (≥ 2) for the first two classifications, the variability of the '70–'74 results is high enough to reduce the significance of the results using the *t*-test. However, when looking at the number of articles describing environmental planning and city and urban planning, and not their variability, it is obvious that these areas have been receiving increasing attention. Observing the energy classification one notes that the number of articles appearing between '65 and '69 [7] is nearly equal to those appearing between '70 and '74 [8]. Is it possible that management scientists have been able to anticipate to some degree the importance of energy planning?

Table 2 presents the frequency results for each of the classifications in *Operational Research Quarterly.*

Significant increases in subject frequency appeared in only 3 or the subject classifica-

[5] It should be noted that the results presented are category dependent. Other classifications could obviously have been selected. Using this set of classifications, 5 graduate students were asked to independently categorize 14 randomly selected articles. With the exception of several misclassifications, their results were virtually identical.

Vol. 6
1978

TABLE 2. CLASSIFICATION OF ARTICLES BY SUBJECT IN *Operational Research Quarterly* FROM 1965 TO 1974

Classification	Year '65	'66	'67	'68	'69	'70	'71	'72	'73	'74	Direction of change and significance
Agriculture, conservation										1	
Bidding				1	1				1		
Communications		2									
Computers	1				1				1		
Construction				1							
Control theory						2					
Contingency planning					1						
Corporate planning					2	1		1		1	
Crime prevention		1				1					
Cutting stock problems										1	
Decision making, decision trees				1	1			1	2		
Design			1			1			1		
Economics				1							
Education	1	1	1	1	1	1	1	2	2		
Energy planning	3		2	1		1	2			2	
Environmental planning	1									1	
Finance	2	1	2	1	1	1	2	1		2	
Fire prevention				1							
Food industry										1	
Forecasting		1	2		4		1	1	3	1	
Functional interfacing (e.g. production–finance)								1			
Garbage collection					1						
Government planning			1							1	
Health planning	1						1	1	1	1	
Information systems					1						
Inventory		1	2	2	5	2	1	2	1	6	
Leather industry									1		
Logistics	1			2	3	1	1	4	6	2	
Manpower planning					2	1	1		2		
Marketing		4	1	3	3	3	2	1	5	1	
Military planning	2			1		1					
Modeling						1					
Multi-criteria problems									1	1	
Networks			1				1	2		2	+,*
OR practice, practitioners, and philosophy		2	2	4	1	1		2	1	2	
OR programming techniques (DP, LP, non-linear programming)	2	2		1	1				2	1	
PERT, CPM		1				1		1			
Probability, statistics	1	1									
Production planning			1	2	2	2		1	2	3	
Productivity	2										
Queuing	1								3	1	
R&D		1	4	1	1	2	1	2		1	
Reliability, replacement, maintenance		1	2	2	1	1		2	3	4	
Resource allocation						2			1	1	+,*
Scheduling, sequencing, assignment	2	4		1	1	1	3	6	3	4	
Simulation	2				1			1	1	1	
Social systems				1							
Systems analysis								1			
Technology										1	
Textile, style goods planning									1		
Transportation, traffic planning		3	1		2	3	4	3	1	2	+,*
Transportation problems	2		1						1	2	
Traveling salesman problem				1			1	1			
Underdeveloped countries										1	
Urban planning							1			1	
Vehicle scheduling			2	1	1	1		3			
Total	24	26	26	30	38	31	23	40	46	49	

+ Number of articles in classification from '70 to '74 significantly greater than from '65 to '69.
* Significant at the 0.05 level.

TABLE 3. CLASSIFICATION OF ARTICLES BY SUBJECT IN *Zeitschrift für Operations Research* FROM 1965 TO 1974

Classification	Year '65	'66	'67	'68	'69	'70	'71	'72	'73	'74	Direction of change and significance
Agriculture, conservation			1			1			2		
Auto purchasing				1							
Channel analysis						1		1			
Control theory, practice	1		1			1			1		
Cutting stock problem	1										
Dynamic programming	2									1	
Economics, production functions			3		1			1	1		
Education						1		1		1	+,*
Employment planning									2	1	
Energy planning		1			1					1	
Environmental planning										1	
Forecasting		1								1	
Finance, insurance			2		2		1	2	1		
Functional interfacing (e.g. production–finance)										1	
Game theory	2					1			1		
Graph theory	1			1			1	1		3	
Growth theory		1						1			
Housing							1				
Integer programming								1	1		
Linear programming	1	5			1		3	1	2		
Linear systems							1				
Logistics	1	1						1		1	
Markov models						1				2	
Mathematical theory					1					1	
Military planning				1			1			1	
Minimax		2		1						1	
Mix problem			1			1		2			
Networks		1		2			1	2	2	2	
Non-linear programming	2	1	2	3	3	7	2	1	5	4	
Optimization theory			1	2	4	2					
Parametric optimization								3			
Probability, statistics	2	1	2			1	1			1	
Production, inventory, scheduling	3	1	3	2	1	5	2	4	5	5	+,**
Purchasing										1	
Queuing	1			1			1	4	1		
R&D					1		1				
Ranking procedures				1	2			1			
Reliability, maintenance				4			1	1	3	2	
Risk analysis				1							
Rubber factory								1			
Space travel								1			
Separable programming						1					
Simulation						1				1	
Stochastic processes			1						1		
Stochastic programming				1					1	1	
Systems design								1			
Transportation, traffic planning		1	1	1				3	1		
Transportation problems		1	1					2		1	
Utility theory							2				
Total	17	17	19	22	17	24	19	36	30	34	

+ Number of articles in classification from '70 to '74 significantly greater than from '65 to '69.
** Significant at the 0.01 level.
* Significant at the 0.05 level.

tions, while none showed significant decreases. The increases resulted in networks, transportation and traffic planning and resource allocation. The first two categories also increased in frequency in *M.S.* Observing the results, one notices that the number of articles appearing between '70 and '74 in logistics and scheduling is also increasing, but due to the variability within years, the increase is not found to be statistically significant. Those articles in *O.R.Q.* describing energy planning, like those in *M.S.*, appeared with similar frequency in both time

TABLE 4. NUMBER OF ARTICLES IN *Management Science* DESCRIBING AN IMPLEMENTATION OR UTILIZING ACTUAL DATA

	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74
Articles describing an implementation in a specific environment										
Number	3	6	6	6	4	4	3	9	20	10
% of total	0.04	0.06	0.06	0.09	0.04	0.04	0.02	0.07	0.13	0.08
Articles not describing an implementation but which utilize actual data										
Number	2	2	5	3	5	3	4	2	8	5
% of all non-implementation reporting articles	0.03	0.02	0.05	0.05	0.06	0.03	0.03	0.02	0.06	0.05

periods. Health care planning and manpower planning articles which appear in *M.S.* with increasing frequency are also receiving more attention in *O.R.Q.*, yet the increases are not found significant. The remaining categories in which changes occurred (increases or decreases) in *M.S.* were not found significantly to change in *O.R.Q.*

Increases in subject frequency found in *Z.O.R.* are observed for education and the inclusive category of production, inventory, and scheduling. The tabulated results are found in Table 3.[6]

The lack of significance of change found in many categories is primarily due to the detailed nature of the classifications resulting in relatively few articles in each one. Surprisingly, the number of articles describing areas of current topical interest are relatively few. There are no articles describing health care planning between '65 and '74. In addition the number of articles describing energy and environmental planning are relatively few. On the other hand, networks and non-linear programming articles are appearing with increasing frequency.

The results of the study to determine the number of articles in *M.S.* describing an implementation or using actual data are presented in Table 4.

The results indicate that the number of articles focusing on an implementation is slowly increasing, but that as a percentage of total articles the figure remains less than 15% each year. In view of the concern often expressed [1, 2, 4, 6] that OR has been relatively unsuccessful in its implementation attempts, maybe it is not surprising that the number of reported implementations is so low. When averaged over the past decade only 6% of the articles presented in *M.S.* have focused on implementations (or approx 12% of the 'Applications Series'). When the number which use actual data are added to the number of implementations, still only 10% of the articles are included.

The increased number of implemented articles reported in 1973 results primarily from the 'Management Science in Canada' issue. The largest increase in the nature of implemented articles arises in health care planning and to a lesser extent in other non-profit areas (e.g. municipal planning) while the largest decline in implemented articles is probably in computer usage.

The results suggest that researchers tend not to validate their models using actual data, with only 32 out of 1031 non-implementation reporting articles published in *M.S.* during the period 1965–1974 indicating real data being used. Naturally, this gives rise to the question of the validity of the hundreds of articles illustrated with numerical examples.

Table 5 compiles the articles reporting implementations and those utilizing actual data described in *O.R.Q.*

The number of articles focusing upon an actual implementation is consistently higher in *O.R.Q.* than *M.S.* The resulting percentage of total articles describing actual implementations is thus significantly higher due to the small number of published articles in *O.R.Q.* It is worthwhile noting that the number of implementations has significantly increased since the '65–'67 period. There seems to be no systematic

[6] Although the total number of articles appearing increases from 17 in '65 to 34 in '74, the number of articles per category utilized remains relatively stable (i.e. 1.54 in '65, 1.40 in '70 and 1.54 in '74). The result is similar for the other journals.

TABLE 5. NUMBER OF ARTICLES IN *Operational Research Quarterly* DESCRIBING AN IMPLEMENTATION OR UTILIZING ACTUAL DATA

	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74
Articles describing an implementation in a specific environment										
Number	5	3	5	10	18	9	11	10	16	11
% of total	0.21	0.12	0.19	0.33	0.47	0.29	0.48	0.25	0.35	0.22
Articles not describing an implementation but which utilize actual data										
Number	0	4	3	3	1	1	3	2	3	2
% of all non-implementation reporting articles	0	0.17	0.14	0.15	0.05	0.05	0.14	0.07	0.10	0.05

TABLE 6. NUMBER OF ARTICLES IN *Zeitschrift für Operations Research* DESCRIBING AN IMPLEMENTATION OR UTILIZING ACTUAL DATA

	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74
Articles describing an implementation in a specific environment										
Number	3	0	0	1	2	1	1	3	2	5
% of total	0.18	0	0	0.05	0.12	0.04	0.05	0.08	0.07	0.15
Articles not describing an implementation, but which utilize actual data										
Number	0	1	0	0	1	0	0	1	0	1
% of all non-implementation reporting articles	0	0.06	0	0	0.07	0	0	0.03	0	0.03

change in the subject areas of the reported implementations. Production, logistics and energy planning are those areas in which implementations are most frequent.

The fewest implementations found in the three journals investigated occur in *Z.O.R.*, yet as a per cent of total articles the value differs little from *M.S.* In addition there does not seem to be any noticeable change in the number or subject area of the implementations. Table 6 presents the results.

CONCLUSIONS

There has been a change in emphasis in each of the three journals analyzed, but it is most pronounced in *M.S.* Articles in a number of subject areas have been presented significantly more frequently over the past 5 years, while other subject classifications have been receiving significantly less attention. Such changes, but to a lesser extent, also arise in *O.R.Q.* and *Z.O.R.* The number of implementations described in *M.S.* has remained consistently low, although a perceptible change has been evident during the past few years. As a proportion of total articles, the number of articles describing an implementation is approx 6% during the period '65–'74 (and rises to only 9% during the period '72–'74). The implementation rate is approximately the same in *Z.O.R.*, but substantially higher in *O.R.Q.*

The evidence, when viewed comprehensively, suggests that the literature has either failed to reflect an increasing number of existing implementations or that the number of implementations has remained consistently low. In addition, the results suggest that since fewer than 4% of the non-implementation reporting articles surveyed utilize actual data, one may at least question the validity of the hundreds of remaining models published in the literature.

ACKNOWLEDGEMENT

The authors would like to thank Richard Barton for his valuable computing assistance.

REFERENCES

1. DOKTOR R & HAMILTON W (1973) Cognitive style and the acceptance of management science recommendations. *Mgmt Sci.* **19**, 884–894.
2. HUYSMANS J (1970) *The Implementation of Operational Research.* Wiley, New York.
3. RADNOR M, RUBENSTEIN A & RANSIK D (1970) Implementation in operations research and R&D in government and business organization. *Ops Res.* **18**, 967–991.
4. RUBENSTEIN A, RADNOR M, BAKER N, HEIMAN D & MCCOLLY J (1967) Some organizational factors related to the effectiveness of management science groups in industry. *Mgmt Sci.* **13**, B508–518.
5. SHAKUN M (1972) Management science and management: implementing management science via situational normativism. *Mgmt Sci.* **18**, 367–377.
6. URBAN G (1974) Building models for decision makers. *Interfaces* **4**(3), 1–11.
7. WAGNER H (1972) Commentary on ORSA guidelines. *Mgmt Sci.* **18**(10), B607–613.

ADDRESS FOR CORRESPONDENCE: *Prof. Steven E Permut, School of Organization and Management, Yale University, New Haven, CT 06520, USA.*

ol. 6
978

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 53–58, 1978. Pergamon Press. Printed in Great Britain

Social Responsibility Goal Perception

HARVEY KAHALAS

State University of New York at Albany

ERNEST B UHR

Virginia Polytechnic Institute and State University

(*Received January 1977; in revised form August 1977*)

The intent of this study is to distinguish areas of difference in the perception of organizational goals among managers, hourly workers, and graduate business students. The managers were chosen to represent the present point of view of management, while the hourly workers were selected to be indicative of labor's viewpoint toward the goal selection of organizations. Students were chosen because they represent the potential transfusion into management. The tool used to solicit these responses was a questionnaire composed of twenty questions, with each question divided into two parts. The first part examined the way the individual perceived the achievement level of the organization ('is'), and the other half represented the level that the individual aspired for the organization to achieve in the future ('should be'). These two parts were quantified using a Likert scale ranging from 1 (of no importance) to 5 (of extremely high importance). The results indicate the similarities and differences between each of these three groups in terms of present perceptions and future expectations of these goals.

SOCIAL RESPONSIBILITY GOAL PERCEPTION

FOR MANY years, organizational goals have been an area of interest and study for organization theorists. Within recent years, added attention has been given to the specific external goals as well as the internal goals of the business corporation. In viewing these goals, some writers have examined the various clients and future client groups of the business institution. An interesting examination of these individuals deals with their perceptions of the existing goals of the business firm. An additional step would include an examination of what the organizational goals should be, as perceived by these same client groups.

PERCEPTION

Perception of ideas, development and actions is a complicated cognitive process that is subject to the internal actions of an individual's mind. Additionally, "the cognitive map of the individual is not a photographic representation of the physical world; it is, rather, a partial, personal construction in which certain objects, selected out by the individual for a major role, are perceived in an individual manner" [4, p. 20]. Furthermore, the perception process is not only a selective cognitive process by the individual, but also its results may differ significantly from reality. Nevertheless, since this process is viewed by the individual as reality, his actions and concepts will, to some extent, be based upon these factors. By aggregating the perceptual view of organizational goals by client groups, the hope would be that substantially deviant individuals would not significantly influence the results.

One major factor which has an impact on individual perception is the particular organization and the level within that institution where the person is located [3]. This is particu-

larly important in the business firm where the superior and subordinate may very well have totally different perceptions of the environment [5, p. 355]. In fact, "... the perceptual world of the subordinate is quite different from the perceptual world of the superior, and both may be very different from reality" [6, p. 336]. In essence, individuals can perceive the same event in a different light from that of other persons observing the same event, and the perceptions of individuals or groups at one level may be different from individuals at other levels within or between organizations.

ORGANIZATIONAL GOALS

One organizational factor which is part of an employee or future employee's perception is the goals of the employee's organization. According to Perrow, "Organizations are tools to achieve various goals. To understand them fully, one must understand the goals they pursue" [7, p. 133]. Cyert and March view organizations as coalitions of individuals [1, p. 27]. These individuals range from shareholders to members of society at large, and each has goals of his own. To reduce these individual goals into a common goal or goals, there must be some mechanism to facilitate the arrival at a common denominator. This device is bargaining within and between the coalitions. Further, this bargaining process is depicted as lying on a continuum from the entrepreneurial goal setting to consensual goal setting. Neither represents the real world solution, which lies somewhere in the gray area between the two extremes.

As experience increases within a group coalition, demands calling for bargaining arise which change the texture of organizational goals. This bargaining has lead to a vast divergence of individuals' perceptions of organizational goals, which, in turn, has lead to an even louder disagreement in bargaining within and between coalitions. At the individual level, people's perceptions are changing to the point of view that organizations have a social obligation. At the same time, organizations claim that they either do not have the necessary resources or have no obligation to be socially responsible. There seem to be distinct indications that although organizations are reluctant, a trend is building toward a shift to socially responsible actions by organizations [2, p. 321].

In examining Cyert and March's [1, pp. 39–43] and Perrow's [7, pp. 135–136] goal typologies, there seems to be some difficulty in attempting to quantify these goals. This is basically due to the inclusion of internally and externally generated goals under one broad heading. Because of this problem, it may be more beneficial to examine initially specific goals using a social (external) vs. organization (internal) dichotomy as the only classification device.

Goals can no longer be depicted as broad generalizations, but must be more explicitly stated and identified to meet the internal and external demands made by the environmental coalitions. Although the typology of internal and external goals remains at the broad generalized level, the coalitions can at least recognize the area towards which action is directed while the subgoal depicts the explicit area toward which action is directed. This should aid organizations in their ability to communicate their goals to their internal and external environments.

Numerous educational, social, and economic circumstances have arisen that have made it necessary for many businesses in America to reach clear, and often new, understandings about their goals. During the late 1960s, there were new demands for businesses to assume new roles and serve new interests. Now, in the 1970s, a widespread financial crisis is making it imperative for businesses to specify the objectives to which limited resources may be directed.

Our work in this paper was developed as a tool to help the private sector delineate goals and establish priorities among them. The instrument does not tell us what to do in order to reach the goals. Instead, it provides a means by which many individuals and constituent groups can contribute their thinking about desired institutional goals. Summaries of the results of this thinking then provide a basis for reasoned deliberations toward attempting to define goals.

SAMPLE, METHODOLOGY, AND RESULTS

In the present investigation, the population under study consisted of three basic groups. These three groups consisted of production

Vol. 6
1978

managers and hourly workers in the engine component industry in the Commonwealth of Virginia, and MBA students in a university in Virginia. The members of each group were randomly selected. The first group consisted of thirty production managers who had hourly paid workers reporting to them. In effect, they held lower level managerial (foremen) positions. All the subjects had been in these positions for a minimum of 5 years. The second group was composed of thirty hourly workers who reported to the above mentioned managers. All had worked in their positions for at least 5 years. Finally, the last group comprised of thirty MBA students currently enrolled in an educational program. These students were randomly selected from the group completing the graduate program that year. The questionnaire consisted of twenty questions dealing with possible organizational goals. The questions were designed to reflect perceptions about two types of goals. The internal goals consisted of those questions which attempted to examine direct impact upon employees. These questions ranged from general job satisfaction to those that attempted to ensure congruency between individual needs and organizational goals. The external goals were designed to examine the impact of the firm upon the society. These ranged from maximizing profits for stockholders to supporting community activities.

Additionally, each question was dichotomized into 'is' and 'should be' factors. The former was designed to gain an understanding of perception about existing goals and the latter to provide information about desired institutional roles.

The responses to the questions were on a Likert-type scale ranging from 'no importance' or 'not applicable' (1) to of 'extremely high importance' (5). The subjects were asked to consider the business institution as a whole and not just their own firm. Furthermore, in giving 'should be' responses, they were instructed not to have their beliefs restrained by whether or not the goal, realistically, could ever be attained.

Two levels of analysis were used for this investigation since there was interest in exploring differences between the three groups of respondents on an overall basis and on a question by question basis. It was felt that this approach would yield greater interpretation possibilities and possibly provide increased insight into the situation under study.

For the overall analysis, the questions were divided into internal and external sets as previously discussed. Individual respondent scores for 'internal is,' 'internal should be,' 'external is,' and 'external should be' question sets were determined by summing the appropriate individual question scores. These summed overall scores were then subjected to the Kruskal-Wallis one-way analysis of variance by ranks test to determine if the differences among the samples signify genuine population differences [8, pp. 184–193]. This test is especially appropriate in this situation since the two assumptions of the test are that the variables under study have an underlying continuous distribution and that at least an ordinal measurement has been used to quantify the variables.

A chi-square test was used for the individual question analysis to determine if the observed differences among samples signify population differences. When using this test, care was taken to insure that the minimum expected cell size was five, as required.

In examining the results of the overall analysis of variance tests, the internal 'is' ($H = 15.14$, $P > 0.001$) and 'should be' ($H = 22.44$, $P > 0.1001$) questions were statistically significant at the 0.10 level. The external 'is' ($H = 1.87$, $P \cong 0.4$) summation was not statistically significant at the 0.10 level. The external 'should be' ($H = 8.74$, $P \cong 0.013$) was statistically significant at the 0.10 probability level.

When the analysis was reduced to the individual questions, a number of variables were identified as statistically significant at the 0.10 level. For the internal questions, the eight statistically significant 'is' responses were numbers 1, 7, 9, 12, 16, 17, 18, 19, while the nine statistically significant 'should be' responses were questions 6, 7, 10, 11, 12, 13, 16, 17, 18. The only external question showing any statistically significant difference between the three various groups was question 5, 'should be.'

The managers basically viewed the statistically significant 'is' questions as of medium importance. The exceptions were questions 1, 7, 12, 17 and 19, which were viewed as medium to high importance and question 18, which was evenly spread across all possible responses.

TABLE 1. QUESTIONS OF GOAL PERCEPTION

	Question
1.*	to help employees gain job satisfaction.
2.†	to maximize profits for stockholder.
3.*	to help employees develop a sense of self-worth and self-confidence.
4.†	to provide equal opportunity employment for all members of the society.
5.†	to provide training for the disadvantaged members of the society.
6.*	to provide recreational facilities for employees and their families.
7.*	to provide method of improving job skills through continuing education.
8.†	to minimize cost and pass saving on to consumers.
9.*	to create an atmosphere where organizational goals and individual needs can be in harmony.
10.*	to provide a safe and healthy environment for the employees.
11.*	to promote growth to assure continuous employment for their employees.
12.*	to encourage employee creativity and innovation.
13.*	to upgrade present working conditions.
14.†	to conform to government regulations.
15.†	to actively support community activities.
16.*	to punish employees who fail to perform effectively and efficiently.
17.*	to allow employees to participate in decisions that directly affect their day-to-day job.
18.*	to search for remedies to job frustration.
19.*	to make promotions based on ability.
20.*	to keep employees informed of their job performance.

* Internal type goal.
† External type goal.

For all the statistically significant internal 'is' questions, hourly workers ranked the response as of low or no importance. The only exception was question 16, where the response was dichotomized between low importance and high importance.

The MBAs' responses to the statistically significant internal 'is' questions were quite diverse. On questions 7, 16 and 19, the responses were medium to high importance, numbers 12 and 17 were viewed as low or no importance, the responses to questions 1 and 18 were evenly distributed, and question 9 was viewed as medium to low importance.

When examining the results of the statistically significant internal 'should be' responses, hourly workers responded with high importance for all questions. These responses ranged from high for most responses to a question to almost complete agreement among the subjects for a high level response.

The responses of the managers were more diverse. Questions 6 and 7 were ranked medium to low importance, questions 11, 13 and 17 of high to medium importance, and numbers 10, 12, 16 and 18 of high importance.

The only statistically significant external question was number 5, 'should be,' which was ranked medium to low importance for managers and MBAs and high importance for hourly workers.

When viewing the total internal responses, questions 7, 12, 17 and 18 were the only questions statistically significant on both the 'is' and 'should be' factors. Additionally, other trends existed with the significant internal 'is' questions. Here, the hourly workers never agreed with the managers, although on certain questions, the responses were similar to MBAs or were not similar to either of the other two groups. With the significant internal 'should be' questions, one trend which was identified was the lack of agreement between the MBAs and hourly workers. At no time did the results of these two groups agree on any significant question, although all other combinations could be identified. These two specific trends dealt with a lack of agreement between hourly workers and another group.

IMPLICATIONS

The implications of this study are quite significant for those studying organization goals. There are two factors which are observed immediately from the data analysis. The first deals with the almost nonexistent differences in the external questions. Only one question (question 5) was identified as being significantly different in terms of group perception of an organizational goal, and that existed only with the 'should be' portion. Basically, therefore,

these groups may be viewed as not perceiving any significant difference in the external dimension. For our sample of three groups—managers, hourly workers, and MBAs—they jointly perceive the business corporation operating in the same manner in this area. Additionally, each of the groups desires to have the corporation operating in the same way.

Cyert and March suggest that to combine individual goals into common goals, bargaining within and between the various coalitions must occur. With the identified external goals being almost totally similar between managers and hourly workers, the needed bargaining may be perceived as having already occurred. The only factor which may not allow for this interpretation is the additional similarity of the MBAs. The three group similarity tends to suggest that these factors were bargained at a higher level in the society creating similar perceptions and values in various societal members. This agrees with previous suggestions that existing and desired organizational goals are, to some extent, determined by environmental pressures and social values.

The second area of concern deals with the lack of similarity in the internal goals. This suggests areas beyond existing social values or an area where the society has established a zone of indifference. In effect, these factors are presently ones which the society is allowing each individual or organization to bargain according to their own desires and perceptions. Additionally, each group may be operating consistently with its own subgroup's values. This is a particularly likely interpretation if we accept the relationship between an overall system establishing value boundaries within which each subsystem or subgroup must operate. This approach only limits the subsystem's discretion only to a zone of decision making and not to a specific response to some organization goals.

With these internal type questions, one factor concerns the dissimilarity between hourly workers and managers ('is') and MBAs ('should be'). The hourly workers never agree with managers on the statistically significant 'is' questions and never desire organizations to operate in the same manner as desired by MBAs. Therefore, the 'is' portion may be identified as a perception factor with hourly workers and managers never perceiving the operation of an organization in the same manner. Furthermore, the 'should be' section of the question may be viewed as an expectation concept with hourly workers and MBAs' never agreeing on organizational expectations.

Finally, when examining the four questions dealing with internal goals which indicate statistically significant differences between the three groups on both the 'is' as well as the 'should be' responses, (questions 7, 12, 17 and 18), three basic points emerge. The most obvious concerns the desire by all groups that the corporation should view each of these factors with greater importance. Additionally, with these four questions, the managers view the corporation as operating on a higher level than the other two groups for the 'is' question while the hourly workers are the highest group for the 'should be' questions. This would tend to reinforce existing theories that managers view the corporation as doing a great deal in many job-related social responsibility areas such as continuing education for workers, allowing participation in decision making, and working to solve job frustration. On the other hand, hourly workers feel that these corporate responsibility factors should be at a level higher than the expectations of the other two groups.

The findings of this study are a preliminary attempt to examine a number of factors simultaneously. The examination viewed three different groups and identified their perception and expectations with respect to organizational goals of both a market or traditional nature as well as those of a social or corporate responsibility complexion. Accordingly, the results show the similarities and differences with regard to perceptions and expectations of these goals and their importance in terms of desires by the groups. Additionally, they provide a basis for future studies in perception of organizational goals. Future studies should deal with expanding the sample size and refining the groups and types of questions to gain a better perspective of goals, perceptions, and social responsibility.

REFERENCES

1. Cyert RM & March JG (1963) *A Behavioral Theory of the Firm.* Prentice-Hall, Englewood Cliffs.
2. Davis K (1973) The case for and against business assumption of social responsibilities. *Acad. Mgmt Jl* **16**, 312–322.

3. DEABORNE DC & SIMON HA (1958) Selective perception: a note on the departmental identification of executives. *Sociometry* **21**, 140–144.
4. KRECH D, CRUTCHFIELD RS & BALLACHEY EL (1962) *Individual in Society*. McGraw-Hill, New York.
5. LEAVITT HJ (1950) *Managerial Psychology*. University of Chicago Press, Chicago.
6. LUTHANS F (1973) *Organizational Behavior*. McGraw-Hill, New York.
7. PERROW C (1970) *Organizational Analysis: A Sociological View*. Wadsworth, Belmont.
8. SIEGEL S (1956) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

ADDRESS FOR CORRESPONDENCE: *Professor Harvey Kahalas, School of Business Administration, SUNY-Albany, 1400 Washington Avenue, Albany, New York 12222, USA.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 59-64, 1978. Pergamon Press. Printed in Great Britain

Market Share Model of Rural Telephone Service

JG CHURCH

IM GORDON

Bell Canada

(*Received March 1977*)

Premium telephone rates, population density, and the grade of telephone service that rural customers choose all vary with distance from the nearest urban center. A survey of 3800 rural subscribers determined their choice of service for alternative prices given the addition of a new service. A two-stage probit model was used to estimate the new market shares for a wide range of conditions. The model was applied, as part of a large-scale revenue/cost simulation system, to the current and forecasted distribution of customers to evaluate the economic impact of numerous pricing and provisioning possibilities.

INTRODUCTION

BELL CANADA provides telephone service to the majority of residents in the Canadian provinces of Ontario and Quebec. Its territory is divided into approx. 1200 geographic entities called switching centres and includes about 420,000 rural customers. A typical rural switching centre includes a town or small city plus the surrounding rural territory.

The basic telephone service for customers living outside the boundary of the urban centre is multi-party service where up to 10 customers share the same line. These rural customers may face three forms of inconvenience: they may hear the ringing of calls to other customers on the line; they may find the line unavailable for their incoming or outgoing calls; and they may find that others listen in on their calls.

Urban grades of service (1 and 2 party) are also available to rural customers in many instances; approx. 40% have 1 or 2 party service. To compensate partially for the large capital expenditure required to provide 1 and 2 party service in areas of low population density, rural customers pay more for these services than customers within the urban boundary. The amount increases linearly with the distance between their residence and the boundary, as shown in Fig. 1.

In 1974 there were 29,000 unfilled requests for 1 or 2 party service. This unfilled demand, coincident with a request from the regulatory body to investigate rural service, led Bell Canada to study economical means of improving rural telephone service. Providing a 4 party service option to all customers outside the urban boundary at a price close to multi-party service with no mileage charge was seriously considered. It would offer increased convenience at a lower monthly cost to the customer and lower capital cost to the telephone company than urban grade service. Since the pricing structure for 4 party service might not cover the associated costs, consideration was

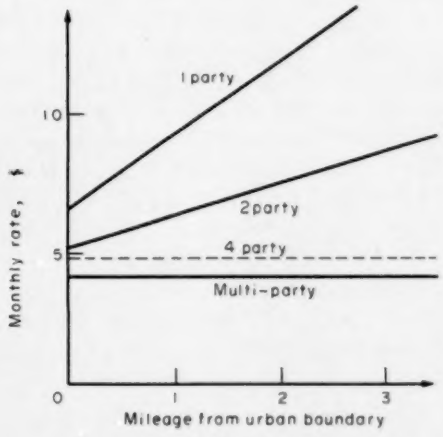


FIG. 1. *Example of rural service pricing.*

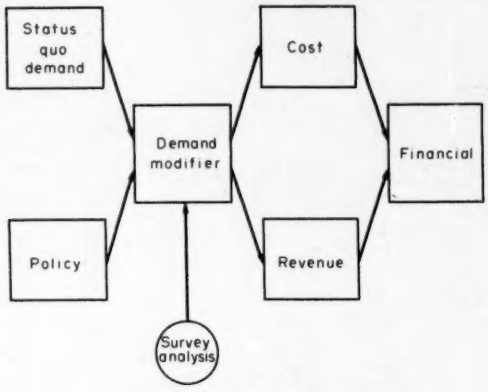


FIG. 2. *Rural service planning model.*

given to raising the mileage charge for rural 1 and 2 party service. This might also reduce demand for the more costly 1 and 2 party services.

Introducing 4 party service might involve capital expenditures of hundreds of millions of dollars, a 4 year phased introduction, and a time span of up to 15 years for revenue and cost evaluation. Policy decisions included the pricing of the services, whether to phase out or continue multi-party service, etc. The size of the investment involved and the complexity of the problem warranted an intensive evaluation of the alternatives. To meet this need, a large-scale rural telephone service evaluation simulator was developed.

The structure of the evaluation simulator can be summarized as follows (see Fig. 2). Forecasts of demand for each of the three existing grades of service under status quo conditions of price, etc. were made for each of the up to 28 quarter-mile bands surrounding each of the 1200 switching centres for the 15 year planning period, i.e. over 500,000 individual items. Policy was set by year and switching centre on changes in services, prices, etc. Information from both sources was input to the demand modifier module for each combination of conditions and the module estimated the revised demand for each grade of service. After this, the simulator estimated the incremental discounted capital costs, expense and revenues and determined the rate of return, net present value, etc., resulting from the policy. This paper will focus on the development and calibration of the demand modifier module which was used to estimate the market share for each of the services under new conditions.

To provide data, a survey of a representative sample of 3800 rural Bell Canada subscribers was conducted. Each respondent was told the monthly rate for his current grade of service and asked to indicate both first and second choice under a scenario where 4 party service was available and new prices applied to 1 and 2 party service.

Data that were available in the simulator as input to the demand modifier module were:

- the number of customers using each of the three grades under status quo conditions;
- the current service prices;
- the new choice of services and their new prices;
- demographic information (province, rate group, and quarter mileage band) defining the 'cell' of customers;
- the number of customers per multi-party line, i.e. line load.

The demand modifier estimates of change in customer service preference were restricted to this set of potential explanatory variables.

METHOD OF ANALYSIS

Estimates of the customer preferences for the four possible grades of service for each combination of conditions were based on data gathered at the individual household level. Behavioural models of individual choice based on a probabilistic mechanism were developed in psychology [5] and have been used extensively in modal split aspects of urban transportation planning [10]. The three main approaches of this type are discriminant, probit, and logit analysis.

The discriminant choice model has been used by several researchers [6, 7, 9] but has been found by Stopher and Lavender [7] to be inferior to probit and logit models in reproducing original sample data. Probit and logit models are based on the disutility of variables measured at the individual level but differ in their assumptions on the underlying distributions [3, 4, 8]. Stopher and Lavender found no significant difference between logit and probit predictive abilities. Since the appropriate computer routines were available, the probit approach was selected.[1]

[1] Probit routines were made available through the courtesy of Prof. M. Dagenais, L'Universite de Montreal.

Vol. 6
1978

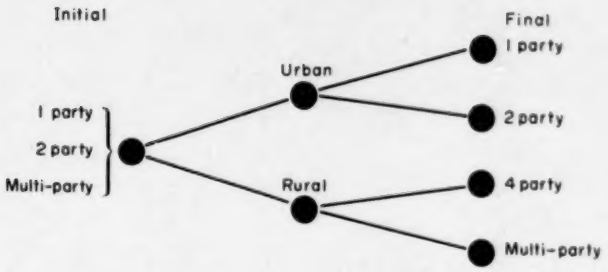


FIG. 3. *Sequential probit model.*

MODELLING

The market share modelling problem is one of estimating the transitional probability of a customer moving from any of the three initial states (1, 2, multi-party) to the four possible final states (1, 2, 4, multi-party) as a function of the various explanatory variables.

Multinomial probit analysis was felt to be not sufficiently reliable to estimate models involving choices with more than two alternatives [10]. Watson [10] also suggested that a series of binary decisions might better represent the actual choice process when multiple alternatives are involved. As a result, the grade of service choice process was decomposed into two sequential binary splits.

Grades of service were grouped into urban grades (1 and 2 party) and rural grades (4 and multi-party). Previous internal studies of rural customers had shown a reasonably well-defined separation of customers into groups with needs corresponding to urban and rural services. Using this natural division, the customers were first split between urban and rural service, and then split further within urban and rural grades, (see Fig. 3).

A large number of combinations of the available variables were fitted to survey data using the probit model and tested to determine their importance in explaining consumer response to changes.

VALIDATION

The probit models were developed and validated using the three following criteria. Model coefficients were retained on the basis of *t*-ratio significance tests. The significance of the overall models were subjected to likelihood ratio tests (for a detailed explanation of this, see [2]). The predictive ability of each model was compared with the actual survey results. The models produced final state patterns that showed minimal deviation from the survey results.

RESULTS

The survey data included information on customer response for numerous combinations of the explanatory variables. The probit-models quantified the relationships embedded in these cross-sectional data to provide estimates of market share at a level of disaggregation and for combinations of conditions that were not possible using the survey results directly. The independent variables entered different probit models with differing degrees of importance. They were, in decreasing order of importance:

- existing grade of service;
- multi-party line load;
- price;
- demographics.

Each of these variables is discussed in more detail below.

Grade of service

The customer's current grade of service is the single strongest predictor of the new grade preference. For example, over half of the multi-party customers would retain their service even if 4 party service were offered at a rate only 10% higher than the current multi-party rate (see Fig. 4). The relationship between current grade and new grade selection was particularly important in estimating the market share under new conditions in the future. Estimates

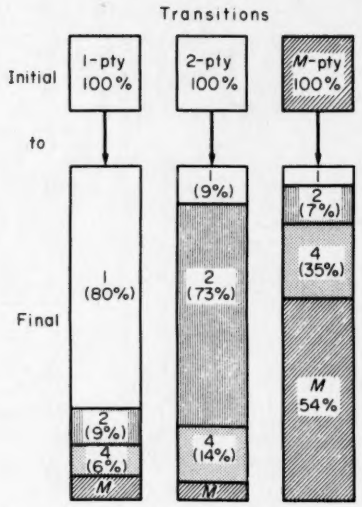


FIG. 4. *Grade of service effect (average 1 and 2 party price increases of 30%).*

of the grade of service distribution by year under status quo conditions were based on historical information. The probit models were applied to these data to generate future new status market share estimates.

Multi-party line load

The number of customers sharing a multi-party line varied from 3 to 10 depending on the facilities available. The data confirmed the a priori assumption that higher line load customers would find 4 party service more attractive. Since lowering line load would require extensive new facilities and could also affect the relative attractiveness of 4 party service, line load was investigated as an important policy variable. For example, decreasing the load from 10 to 3 per line had more impact on customer choice than changing the 4 party price premium from $1.00 to zero (See Fig. 5). The analysis showed that line load had a larger effect than the 4 party to multi-party price differential.

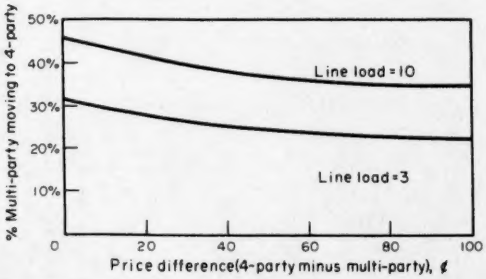


FIG. 5. *Multi-party response to price and line load.*

Price

Price is often considered to have the most important influence on demand. In this study the effect of price was statistically significant in all of the choice models. However, the explanatory power of prices varied with the current grade of service and the particular choice situation. Price also appeared to be relatively less important than other factors such as line load.

There were three important price differences that affected the binary probit splits. The price differential between 2 and 4 party service affected the choice between urban and rural type services for all customers. The price difference between 4 and multi-party services affected the choice between these grades only for current multi-party customers; this may reflect the relative disinterest in these services shown by current urban-grade customers. The price difference between 1 and 2 party service affected the choice between them only for current urban grade customers, perhaps for similar reasons.

Contributing to the reduced impact of price on the final distribution of customers are the peculiarities of the rural telephone service pricing structure. For example, no price changes were considered for either the 1 and 2 party service price applying at the urban boundary or for the multi-party price. Changes were limited to the price for the new 4 party service offering and the additional charge per quarter mile applying to 1 and 2 party service. To maintain a reasonable relationship between services, the 4 party price was also restricted to fall between the basic 2 party rate and the multi-party rate.

Demographics

Available demographics were the customer's province, urban centre size, and mileage from the urban boundary. Province is a good indicator of the different cultural and linguistic backgrounds of the population: Quebec is predominantly French-speaking while Ontario is mainly English-speaking. However, province appeared to have a weak and inconsistent relationship to customer preferences.

Urban centre size and mileage from it can be viewed as indicators of 'ruralness'. Although the percentage of customers currently having

each grade of service varied by urban centre size, no relationship was found between size and customer preference. Perhaps the potential contribution of centre size had already been reflected in the current distribution of grades of service. There was a weak effect of mileage, indicating the customers farther away from the urban boundary were less likely to choose the higher grade of service.

EXTENSIONS OF THE MODEL

The market share model was extended to several potential scenarios that were not directly included in the survey. One of these involved the possibility of lowering multi-party line load without introducing 4 party service. The survey data on the customers' second choice of grade of service were used to model situations where one or more services were not available. Customers who had initially preferred 1, 2, or multi-party service were not affected by the change and were assigned to their most preferred choices. Those who initially preferred 4 party were reallocated using new choice models based on their second choice preferences.

Very few 1 or 2 party customers made 4 party their first choice so that these customers were reallocated to 2 and multi-party service using the simple proportions identified in the survey. However, the majority of multi-party customers' first preference was multi-party and many did not make 4 party their second choice. The reallocation of these customers was sufficiently complex that another set of binary probit models was developed: one to split them between urban and rural services, and one between 1 and 2 party service.

CONCLUSIONS

The large-scale simulation of rural telephone service that was developed by Bell Canada estimated the revenue and cost implications of providing 4 party service under alternative scenarios. A number of the more important options that were considered, with a broad range of 1 and 2 party mileage rates, include:

- phasing in 4 party service with and without retention of multi-party;
- not introducing 4 party but reducing multi-party line loads to 4;
- no 4 or multi-party service.

The results generated by the evaluation model formed the basis of the Bell Canada submission to the government regulatory body [11] on the future plans for rural telephone service. The essential elements of the recommended plan were to gradually reduce multi-party line loads to a maximum of 4 per line and then to institute a 4 party charge at 20% to 30% above the current multi-party rate. 1 and 2 party mileage charges would also be changed so that the price was increased an average of 30% for customers within 2 miles of the boundary and held constant at the 2 mile price for greater distances. The changes were to be phased in by switching centre over 4 years and require an estimated additional expenditure of over $300 million. Even with these higher rates, the proposal would have resulted in an estimated after tax net present value shortfall of over $20 million. This original submission was under consideration by the regulators pending final decision and was subsequently revised as part of a general application for a price increase.

The demand modifier module described in this report was the essential element that enabled the simulator to estimate customer response to a wide range of potential situations and hence the financial implications of alternatives involving many millions of dollars and over 400,000 rural customers.

REFERENCES

1. Hansen S (1974) Structure of disaggregate behavioral choice models. *Behavioral Demand Modelling and Valuation of Travel Time*, Special Report 149, Transportation Research Board.
2. Kendall MG & Stuart A (1968) *The Advanced Theory of Statistics*, Ed. 2, Vol. 3. Hafner, New York.
3. Lave CA (1969) A behavioural approach to modal split forecasting. *Transpn Res.* **3**, pp. 463–80.
4. Lisco TE (1967) The value of commuters' travel time: a study in urban transportation. *Ph.D. Dissertation*, Department of Economics, University of Chicago.
5. Luce D (1959) *Individual Choice Behavior*. Wiley, New York.
6. Stopher PR (1969) A probability model of travel mode choice for the journey to work. *Highway Research Record*, No. 283, Highway Research Board, Washington, DC.
7. Stopher PR & Lavender JO (1972) Disaggregate, behavioural travel demand models: empirical test of three hypotheses. *Proceedings, Transportation Research Forum*, pp. 321–36. Oxford, IN.
8. Talvitie A (1972) Comparison of probabilistic modal-choice models: estimation methods and system inputs.

Highway Research Record, No. 392, Highway Research Board, Washington, DC.

9. WARNER SL (1962) *Stochastic Choice of Mode in Urban Travel: A Study of Binary Choice*. North-western University Press, Evanston, Illinois.
10. WATSON PL (1973) A composite model of intra-city mode choice. *Proceedings of the International Conference on Transportation Research*, Transportation Research Forum, pp. 637–41.
11. *Non Urban Service Improvement Comprehensive Programme* (1975) Bell Canada submission to the Canadian Transport Commission.

ADDRESS FOR CORRESPONDENCE: *JG Church Esq, Director Marketing Research and Planning, Bell Canada, 25 Eddy Street, Hull, Quebec, Canada J8Y 6N4.*

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 65–69, 1978. Pergamon Press. Printed in Great Britain

Redeployment and Diversification at Harwell

PHILIP GUMMETT

MICHAEL GIBBONS

University of Manchester, UK

(*Received February 1977; in revised form June 1977*)

Between 1967 and 1974 the United Kingdom Atomic Energy Research Establishment, Harwell, underwent a major transformation. Faced with a declining workload under the atomic energy programme, the laboratory has succeeded, not without difficulties, in defining and implementing a new mission to assist British industry to become more competitive. Within a declining staff total, the proportion of staff employed on this new mission, under contract to government departments and industrial firms, has risen to about 50%.

THE 1960s and early 1970s were times of change for the research establishments of the British government. The demise of the Department of Scientific and Industrial Research in 1964, and the transfer of most of its laboratories to the Ministry of Technology, meant that for the first time these laboratories were accountable to an executive department of state [8, 15, 17]. Some laboratories (such as the Road Research Laboratory and the Building Research Station) were later transferred to other departments while, with the fusion of the Ministries of Technology and Aviation in 1967, all the aviation research establishments came under a civil department [3]. This arrangement changed again in 1970 with the establishment of the Department of Trade and Industry and, soon afterwards, of the Procurement Executive of the Ministry of Defence. A proposal to set up a British Research and Development Corporation, comprising several of the biggest government laboratories, and operating as a contract laboratory outside the civil service, came and went in 1970 [10]. The separation of a Department of Energy from Trade and Industry in 1974 meant the transfer of the Atomic Energy Authority to its third executive ministry in 10 years. In addition, the Rothschild report of 1971 [1] had repercussions over the next few years not only for the research councils [18, 19] but also for the organisations through which some government departments formulate their research policies [4, 6].

Throughout this period of uncertainty and flux the Atomic Energy Research Establishment, Harwell, steered a fairly steady course towards a new role for itself. Harwell has successfully engaged upon a major exercise in redeployment of staff and diversification of activities which has had considerable implications for its relations with industry, and for the attitudes and values of its staff. The stimulus to these changes was the question, faced from time to time by all mission-oriented laboratories, of what would become of the laboratory once its mission had been accomplished.

Redeployment and diversification are not new problems for Harwell. The laboratory was set up in 1946 as part of Britain's nuclear weapons programme [5, Ch. 18]. Increasingly, however, Harwell became responsible for research and development on civil nuclear power, growing to 6000 staff by 1956 [16, p. 56]. After the Suez crisis the reactor programme expanded, and most of the Reactor Division of the Atomic Energy Authority was transferred from Harwell to Winfrith in 1958–60. A Research Group of the Authority was formally established in 1961, comprising the laboratories at Harwell, Culham (fusion work) and Amersham

[1] This paper is based on a report, *Reployment in British Government Research Establishments*, submitted by the authors to the Science Council of Canada in August, 1975.

(radio-isotopes). Weapons, Reactor, Production and Engineering Groups were also established. This reorganisation, however, coincided with the end of the period of rapid growth for Harwell, and the start of a period of fixed, and later declining, staff. The composition of the staff had also changed throughout this period as a function of Harwell's changing programme; whereas from 1953 to 1959 physicists predominated, by 1963 chemists and metallurgists were more prominent [16, p. 60].

By 1965, with the start of the programme for the second generation of nuclear power stations, it was clear that a watershed had been reached. Nuclear power was now well established as a means of generating electricity, and although much research would continue to be needed for the third generation of reactors, the scale of effort would be reduced to about one-third of the then current level [8, p. 75, and 7]. It was felt within Harwell that to reduce the laboratory by such a large amount would completely destroy its viability, and therefore two objectives were set by the staff. As Dr. Marshall, Harwell's Director from 1968 to late 1975, explained:

> "We needed to reduce the size of the laboratory and therefore our call on atomic energy funds but at a rate which was modest enough to retain the morale and coherence of our research teams and we needed to introduce an additional mission to give renewed vigour and opportunities in the future [8, p. 75]".

The need for a new mission also affected other Atomic Energy Authority establishments, including the Atomic Weapons Research Establishment, Aldermaston. It was largely with an eye to such problems that the Science and Technology Act (1965) made provision, in Section IV, for the Authority to undertake research "in such matters not connected with atomic energy as may, after consultation with the Authority, be required by the Minister of Technology", the Authority by now being accountable to that minister. Within months of the Act being passed the Authority noted that a wide variety of non-nuclear projects was under consideration, including one on desalination at Harwell [14, p. 62].

At the same time, Harwell was considering what its new mission should be. According to Dr. Marshall, there were no great difficulties here. Referring to "the constant complaint that British scientists regularly produce good scientific ideas which led fairly rapidly to an improvement in the balance of payments of other countries but not of our own", he continued, "we decided that an additional Harwell mission to help industry with research and innovation was one which was important, attractive to our staff and likely to win Government backing [8, p. 75]". Among the alternative missions which were considered but rejected were that Harwell might become either an advanced research institute, offering to the universities such centralised facilities as a high flux beam reactor (plans for which were then being laid in the British scientific community), or a kind of Technische Hochschule, offering high level teaching and research in certain fields of technology.

THE INDUSTRIAL PROGRAMME

Once the decision had been made, and the ministerial permission obtained, to proceed with the industrial programme (as it came to be called), progress was swift. Between 1966–67 and 1974–75, and within a declining staff total, the proportion of Harwell's staff employed on work for industry and for government outside the atomic energy programme grew from 5%–50%. Over the same period, earnings from the industrial programme grew to about half of Harwell's total income (the remainder coming chiefly from the Atomic Energy Authority's Reactor Programme) [8, pp. 76–79]. Of the income from the industrial programme, about 40% came from industrial customers by the mid-seventies, the rest coming from governmental sources for work deemed to be of industrial importance [8, Fig. 3]. The proportion of income on the industrial programme earned for routine services, as opposed to research and development, fell from about 60% in 1966–67 to about 20% in 1974–75 [8, p. 78].

The industrial programme was organised into fifteen major 'projects', such as atmospheric pollution, desalination, carbon fibres, ceramics, high temperature chemical technology, computer software development, analytical R&D, non-destructive testing, and heat transfer and fluid flow. Each project involves staff from one or more of Harwell's twelve disciplinary-based divisions and may, as with carbon fibres, involve up to six divisions. This matrix pattern of responsibility (disciplinary

Vol. 6 1978

divisions cross-cut by projects) is designed to allow multidisciplinary teams to be assembled on a flexible basis. Each project is supervised by a project manager (typically a Senior Principal Scientific Officer) who is responsible to the Assistant Director in charge of the industrial programme, and who can turn for guidance to his own division head and to Harwell's Marketing Department (see below) [11, pp. 146–147].

The commercial projects undertaken by Harwell have ranged from those fully paid for by a firm, to joint developments with both sides committing money and effort under an agreement to share any benefits. There has been a noticeable tendency for firms to place their first contracts with Harwell on a joint development basis, becoming sufficiently confident of technical success by the second or third contract to opt to accept all the financial risks (and all the benefits) themselves [13, p. 573, question 176]. Related to this tendency is the observation that the various types of facility offered by Harwell have begun to pay their way at different rates of progress. It takes 3–5 years for a project group to establish its reputation in contract research, whereas routine services can begin to pay their way much more quickly [13, p. 599 and 12].

In developing its capacity to conduct commercial work, Harwell has had many difficulties to overcome in getting close enough to firms to be able to see problems through their eyes, in avoiding conflict with work being done in other laboratories, and in convincing customers that it could work within commercial constraints [12, p. 4]. Some help with the second and third of these difficulties derived from its decision to restrict its industrial work to areas in which it was already technically excellent, thus avoiding the need simultaneously to learn new science and business practice [8, p. 76]. The task of getting close to firms can only have gained from Harwell's decision, in sharp contrast with the usual practice in British government research establishments whereby knowledge is made available to all comers, to cooperate in very many cases with single companies and in secret. This "Principle of Maximum Unfairness", as it was dubbed, has been justified on the grounds that companies will be more likely to place contracts with government laboratories under conditions which do not split the market in the early stages of development, thereby giving the firm a chance to establish a lead over its competitors [11, p. 147].

DEVELOPING ENTREPRENEURIAL SKILLS

In redeploying staff on the scale that Harwell has, much depends on their quality. Here Dr. Marshall has observed that while such commercial experience as it had at its disposal has been valuable to Harwell, the most important factor governing success has been the cultivation of the entrepreneurial skills of which, in his view, a good scientist is capable given the right environment and encouragement [8, p. 80]. At Harwell part of that environment has been created by the lead of the senior staff themselves. The creation of the right conditions for the success of the industrial programme has also depended upon the work of Harwell's Marketing and Sales Department. As at the only other British government civil research establishment with a well-developed marketing function, the National Engineering Laboratory, Harwell's Marketing Department does not seek to dominate the marketing process but adopts instead a philosophy of decentralisation. According to Dr. Clarke, then Marketing Director and now Director of Harwell's Energy Programme, care must be taken not to dampen the innovative spirit of the individual scientist by having marketing information thrust upon him from someone whom he does not know as a colleague.

> "This would be highly undesirable. Successful innovation is founded, not in a committee process, but on the motivations and energies of entrepreneurs. The marketing input must inform the judgement of the entrepreneur, not replace it [2, p. 12]".

Harwell, accordingly, has instituted a training scheme under which scientists and engineers learn to operate professionally in the marketing function and then return to their research work. The basic idea, to quote Dr. Clarke again,

> "is to give the technical innovators, the same access to, and the same type of relationship with, the marketing function that they have vis-à-vis the technical function."

Writing in 1974, when this educational programme had been going for only 2 years, Dr. Clarke felt able to claim that it was already

bearing fruit. The scientists and engineers were increasingly building marketing knowledge into their work, with the interesting consequence that some customers were including marketing components in their contracts for research at Harwell.

While the emphasis from the Director downwards on the importance of developing entrepreneurial skills has evidently been crucial to Harwell's successful redeployment, it has, nevertheless, had its difficulties. Chief among these has been that of overcoming the resentment which has been felt in some, limited, quarters that years of solid and painstaking research may now count for little, when it comes to promotion, alongside a record of quick success in securing industrial contracts. The fear that entrepreneurial values may displace the older-established research values does not seem to us to have spread far at Harwell. We have, nevertheless, encountered it and think it worth recording if only as a warning to other laboratories which might someday find themselves following in Harwell's footsteps.

Another difficulty derives from the fact that, because Harwell is a government laboratory, the entrepreneurial instincts of its staff must never be allowed to work against the national interest. The question of what is, or is not, in the national (or public) interest has been a prime subject for political philosophers since at least the seventeenth century. Within the civil industrially-oriented laboratories of the British government, particularly those funded through the research requirements boards of the Department of Industry, it tends to be understood pragmatically in such terms as effect on balance of payments, effect on relationship between sections of industry, effect on competitiveness of British industry, and so on [6]. Put negatively, the main constraints are that great care must be exercised in dealing with foreign, or foreign-owned, firms, and that government research establishments must not directly compete with industry in the provision of research services. These constraints limit the possibility of direct comparison between Harwell and strictly commercial laboratories.

CONCLUSIONS

Harwell's success in defining and moving towards a new role has depended on many factors. Not least of these was the farsighted leadership of the senior staff who, anticipating a decline in work for the Atomic Energy Authority, did not await events but mobilised discussion inside the laboratory and in the Ministry of Technology with a view to identifying a new mission which would be acceptable both to their colleagues and to the ministry. Having found a new mission, Harwell proceeded to diversify from its strengths; that is, rather than expand into new areas of science and technology, it chose to offer its services only in areas where it was already technically competent. It took very seriously the marketing side of its new activities, and has developed a successful marketing philosophy. The derivation of the Principle of Maximum Unfairness demonstrates the commercial astuteness of Harwell's senior staff, and the emphasis on the cultivation of entrepreneurial skills has permeated the laboratory. It has become significant in the promotion of staff, and it continues to be professionally encouraged through the Marketing and Sales Department.

Although smaller than in 1967, Harwell is now about twice the size that it would have been without the industrial programme. Much, therefore, has been achieved in the intervening years. From a wider perspective, however, the value of the endeavour must remain an open question. The public interest may or may not be well-served by preserving Harwell at a size larger than its atomic energy work requires, and allowing it to earn part of its keep through contracts with industrial firms and government departments. Presumably Harwell's customers find its services cost-effective, but the question remains of whether the national interest might not be better served by the dispersal of staff not needed for the atomic energy programme to posts in industry, government and education. This is a question in technological economics which cannot be answered in the present state of the art. In any case, the economic analysis is somewhat beside the point given the political astuteness which Harwell has displayed, and the pressure which the Institution of Professional Civil Servants would undoubtedly generate should a serious attempt be made further to reduce the size of the laboratory. These factors, coupled with the prospect of increasing demand for energy research on a

Vol. 6
1978

broader base than merely atomic energy, suggest a secure future for Harwell.

ACKNOWLEDGEMENTS

We acknowledge the financial support of the Science Council, and also the generous assistance of staff at Harwell and in the Department of Industry.

REFERENCES

1. *A Framework for Government Research and Development* (1971) Cmnd. 4814. HMSO: London.
2. CLARKE F (1974) Three golden rules of contracting. *New Scient.* **62** (supplement), 10–12.
3. CLARKE R (1973) Mintech in retrospect. *Omega* **1**, 25–38, 137–163.
4. GIBBONS M & GUMMETT P (1976) Recent changes in the administration of government research and development in Britain. *Public Admin.* **54**, 247–266.
5. GOWING M (1974) *Independence and Deterrence: Britain and Atomic Energy* 1945–52. Macmillan, London.
6. GUMMETT P & GIBBONS M Government research for industry: recent British developments. *Res. Policy.* In press.
7. MARSHALL W (1972) The industrial research programme at Harwell. *Atom* No. 185.
8. MARSHALL W (1974) Interaction between Government and industry: lessons from Harwell's experience. *Atom* No. 210.
9. MENCHER A G (1975) *Lessons for American Policy-Making from the British Labour Government's* 1964–70 *Experience in Applying Technology to Economic Objectives.* National Science Foundation, Washington.
10. MINISTRY OF TECHNOLOGY (1970) *Industrial Research and Development in Government Laboratories: A New Organisation for the Seventies.* HMSO, London.
11. ROBERTS LEJ (1970) The new industrial programmes at Harwell. *Atom* No. 165.
12. ROBERTS LEJ (1973) Framework of diversification in the United Kingdom Atomic Energy Authority. Mimeographed.
13. SELECT COMMITTEE ON SCIENCE AND TECHNOLOGY (1972) *Research and Development: Minutes of Evidence and Appendices*, session 1971–72, HC 375. HMSO, London.
14. UNITED KINGDOM ATOMIC ENERGY AUTHORITY (1965) *Eleventh Annual Report* 1964–65. HMSO, London.
15. VARCOE I (1974) *Organising for Science in Britain.* Oxford University Press, London.
16. VICK FA (1966) The atomic energy research establishment, Harwell. (Ed. COCKCROFT J) *The Organisation of Research Establishments.* Cambridge University Press, London.
17. VIG NJ (1968) *Science and Technology in British Politics.* Pergamon Press, Oxford.
18. WILLIAMS R (1972) Independence and accountability after Rothschild. *R&D Mgmt.* **2**, 131–135.
19. WILLIAMS R (1973) Some political aspects of the Rothschild affair. *Sci. Stud.* **3**, 31–46.

ADDRESS FOR CORRESPONDENCE: *Dr Philip Gummett, Department of Liberal Studies in Science, The University, Manchester M13 9PL, UK.*

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 71–81, 1978. Pergamon Press. Printed in Great Britain

Regression with Random Coefficients

LW JOHNSON

Macquarie University, New South Wales

(*Received July* 1977)

The first half of the seventies has seen a virtual explosion of interest in the use of random coefficient regression techniques which, for the most part, have been developed or refined by econometricians. However, given the importance of regression analysis as a managerial tool, it is surprising to find scant mention of random coefficient methods in the management science literature. With this in mind, this paper reviews briefly the various random coefficient models that have been developed and provides a reasonably complete bibliography for further reference.

INTRODUCTION

IN THE usual multiple regression framework, it is normally assumed that the parameters to be estimated are fixed across observational units. However, about thirty years ago, Wald made the statement,

> "In some problems it seems reasonable to assume that the regression coefficients $\beta_1,\ldots, \beta_p$ are not constants but chance variables". [77, p. 586].

This was apparently the first suggestion that in some contexts it may be more meaningful to make the assumption that parameters in a model are themselves random variables. Unfortunately, Wald did not suggest a way to estimate such a model. A few years later, Rubin [58] and Hurwicz [30] examined models with random parameters and suggested maximum likelihood (*ML*) methods as a means of estimation. Rubin's model consists of a single linear equation with K parameters, each of which is distributed normally with unknown mean and variance. He derives a set of *ML* equations for this model which seems to be very complex and difficult to solve in practice since numerical methods must be used. Hurwicz looks at a simple system of two equations with one random parameter and its associated likelihood function, also offering no practical solution to the estimation problem presented.

Even though the papers of Wald, Rubin and Hurwicz do not yield useful procedures for estimation, they did bring to the attention of analysts the arbitrary nature of the fixed coefficient assumption. In the well-known econometrics text by Klein, it is noted that in a cross-sectional economic framework,

> "Individuals differ greatly in behaviour, and it may not be possible to obtain observations on a sufficiently large number of variables so that each unit may be considered to behave according to the same structural equation". [38, p. 216].

Yet, the problem of parameters being treated as random variables yielded no published research until the middle of the sixties except for a paper by Theil and Mennes [74]. The work of Nerlove [42], Rao [46, 47], Zellner [81], Fisk [19] and Theil [75] all added to the understanding of the random parameter problem but were either concerned with the interpretation of coefficients in aggregate economic relationships as being considered random [42, 75, 81] or with models of little applicability to business or economics problems [19, 46, 47].

Then, in 1968 Hildreth and Houck [26] (H–H) looked at a model very similar to that of Rubin and suggested several consistent estimators of the means and variances of the random β's. In the same year, a pair of doctoral dissertations [50, 67] were completed which, to some extent, served as launching pads for much of the current work on random parameter estimation.

The refinement of this earlier work, undertaken in the seventies, has been explosive,

resulting in over thirty published works and at least six doctoral disserations.[1] These range from completely applied studies using random coefficient methods to examination of more complex random parameter specifications. However, from a theoretical point of view, most of the statistical properties of the estimators developed have been asymptotic, little attention being given to their small sample properties. The few Monte Carlo studies that have been done only scratch the surface of the problem.

The purpose of this paper is to briefly survey the various types of random parameter models that have been proposed along with related estimation procedures, to examine the statistical properties of the estimators, especially with respect to small-sample properties, and to provide the reader with a fairly comprehensive list of references on random parameter estimation. Of particular interest to the management scientist are a number of studies that have recently appeared [3, 57, 60, 66] concerning random parameter regression in financial and business contexts.

SWAMY TYPE ESTIMATORS

Swamy [67] has analyzed a model with cross-sectional time series (panel) data under the assumption that parameters are random across individual units but fixed over time. His work has essentially focused on a model of the form:

$$\mathbf{y}_i = X_i\boldsymbol{\beta}_i + \mathbf{u}_i \,(i = 1,\ldots, N) \qquad (1)$$

where $\mathbf{y}_i$ and $\mathbf{u}_i$ are $T \times 1$ vectors of observed dependent variables and unobserved random elements respectively, X_i is a $T \times K$ matrix of observations on K non-stochastic regressors and $\boldsymbol{\beta}_i$ is a $K \times 1$ vector of coefficients assumed to be random across individuals, $i = 1,\ldots, N$. Assuming that a suitable set of panel data is available and N, $T > K$, Swamy has derived estimators for the model's parameters under a variety of assumptions about the distributions of the random elements in (1) (i.e. $\mathbf{u}_i$ and $\boldsymbol{\beta}_i$). His simplest set of assumptions may be written as:

$$E(u_i) = 0, \quad E(\mathbf{u}_i\mathbf{u}_j') = \begin{bmatrix} \sigma_{ii}I & \text{if } i = j \\ 0 & \text{if } i \neq j \end{bmatrix}, \qquad (2a)$$

$$E(\boldsymbol{\beta}_i) = \bar{\boldsymbol{\beta}}, \qquad (2b)$$

$$E(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta}_j - \bar{\boldsymbol{\beta}})' = \begin{bmatrix} \Delta & \text{if } i = j \\ 0 & \text{if } i \neq j \end{bmatrix}, \qquad (2c)$$

$$\boldsymbol{\beta}_i \text{ and } \mathbf{u}_j \text{ are independent.} \qquad (2d)$$

Without going into detail [68, pp 312–313], it is easy to show that if one lets $\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\delta}_i$, equation (1) may be compactly written as:

$$\mathbf{y} = X\bar{\boldsymbol{\beta}} + D\boldsymbol{\delta} + \mathbf{u} \qquad (3)$$

where

$$\mathbf{y} = [\mathbf{y}_1', \mathbf{y}'_2,\ldots, \mathbf{y}_N']', \quad X = [X_1', X_2',\ldots, X_N']'$$

$$\boldsymbol{\delta} = [\boldsymbol{\delta}_1', \boldsymbol{\delta}_2',\ldots, \boldsymbol{\delta}_N']', \quad \mathbf{u} = [\mathbf{u}_1', \mathbf{u}_2',\ldots, \mathbf{u}_N']',$$

and D is an $NT \times NT$ block-diagonal matrix with matrices X_1 through X_N on the diagonal and 0's elsewhere.

If assumptions (2a)–(2d) hold, the disturbance vector $D\boldsymbol{\delta} + \mathbf{u}$ has an $NT \times NT$ covariance matrix of the form:

$$V = \begin{bmatrix} X_1\Delta X_1' + \sigma_{11}I & & \ldots & 0 \\ 0 & X_2\Delta X_2' + \sigma_{22}I & \ldots & 0 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & 0 & \ldots & X_N\Delta X_N' + \sigma_{NN}I \end{bmatrix} \qquad (4)$$

The model implies, then, that variances are constant and covariances are zero over time and that there is no auto or serial correlation, conditions that are sometimes difficult to justify in actual economic situations. For example, it may be more realistic to assume that the disturbance terms may be auto regressive. Swamy [69] has examined the model (1) under assumption that:

$$u_{it} = \rho_i u_{it-1} + \epsilon_{it} (0 < |\rho_i| < 1) \qquad (5)$$

[1] See the references at the end of this survey for a fairly complete listing.

Vol. 6
1978

where $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}\ \epsilon_{js}) = \sigma_{ii}$ if $t = s$ and $i = j$, 0 otherwise, ρ_i are fixed parameters, and u_{it} is an element of $\mathbf{u}_i$.

Assumption (5) implies that the second part of (2a) may be written:

$$E(\mathbf{u}_i \mathbf{u}_j') = \sigma_{ii}\Omega_{ii} = \frac{\sigma_{ii}}{1-\rho_i^2}\begin{bmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & \cdots & \rho_i^{T-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & & 1 \end{bmatrix} \tag{6}$$

if $i = j$ and $E(\mathbf{u}_i \mathbf{u}_j') = 0$ if $i \neq j$. Therefore, the covariance matrix (4) is replaced by a similar matrix with $\sigma_{ii}\Omega_{ii}$ in place of $\sigma_{ii}I$ on the diagonal.

An even more general set of assumptions considered by Swamy [70, 71] replaced (6) with

$$E(\mathbf{u}_i' \mathbf{u}_j') = \sigma_{ij}\Omega_{ij} = \frac{\sigma_{ij}}{1-\rho_i\rho_j}\begin{bmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T-1} \\ \rho_j & 1 & \rho_i & \cdots & \rho_i^{T-2} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \rho_j^{T-1} & \rho_j^{T-2} & \rho_j^{T-3} & \cdots & 1 \end{bmatrix} \tag{7}$$

In words, (7) implies that the covariances between individuals may be non-zero, depend on the units in question (i.e. the choice of i and j), and that the serial correlation between a pair of observations in different time periods declines geometrically with the time distance involved.

The model under assumptions (7), (2b)–(2d), may be much more realistic in some situations than under (2a)–(2d) which is simply a special case of the more complex model, but also more difficult to estimate since one must find estimates of not only $\bar{\boldsymbol{\beta}}$, Δ, and σ_{ii}, but also σ_{ij} for $i \neq j$ and the ρ_i. If the sample size is relatively small, this could pose a real problem in estimation.

Swamy [71, pp. 154–56] describes an approximate Aitken type estimator for $\bar{\boldsymbol{\beta}}$ in the more complex model along with estimators for ρ_i, σ_{ij} and Δ, and suggests several alternative estimators for the same model. Although consistency is an optimal asymptotic property, in the small sample situations sometimes encountered by economists or management scientists consistent estimators may not perform well. An examination of the small sample properties of these Swamy-type estimators by means of Monte Carlo studies would be of great interest to the applied statistician, but, unfortunately, little has been done up to now in this area. We will survey just what has been done later in this paper, but first examine a slightly different generalization of the basic model presented above.

Swamy and Mehta [73] present a model with the additional assumption that the $\boldsymbol{\beta}_i$ not only vary across individuals but over time as well, so that

$$\boldsymbol{\beta}_{it} = \bar{\boldsymbol{\beta}} + \boldsymbol{\delta}_i + \boldsymbol{\alpha}_{it} \tag{8}$$

where $\boldsymbol{\delta}_i$ and $\boldsymbol{\alpha}_{it}$ are vectors of unobserved zero mean random disturbances.

Their model may be written

$$\mathbf{y} = X\bar{\boldsymbol{\beta}} + D\boldsymbol{\delta} + D^*\boldsymbol{\alpha} \tag{9}$$

where $\mathbf{y}$, X, $\bar{\boldsymbol{\beta}}$, D and $\boldsymbol{\delta}$ are as before, $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \ldots, \boldsymbol{\alpha}_N']'$

$\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}_{i1}', \boldsymbol{\alpha}_{i2}', \ldots, \boldsymbol{\alpha}_{iT}']'$

$D^* = \text{diag}[D_1, D_2, \ldots, D_N]$ and

$D_i = \text{diag}[\mathbf{X}_{i1}', \mathbf{X}_{i2}', \ldots, \mathbf{X}_{iT}']$, $\mathbf{X}_{it}'$ being T-vectors of observations on the exogenous variables for individual i. Equation (9) differs from (3) in that if one of the X's is a constant (usually 1), the corresponding random coefficient in (9)

is indistinguishable from the random error term $\mathbf{u}$ in (3), but (9) also allows all other coefficients to vary over time. Hsiao [27, 28, 29] also considers a model with cross and time varying coefficients but his formulation is less general than the one described above, and so will not be discussed here.

Swamy and Mehta develop consistent estimators for the parameters in (9) under assumptions similar to (2) and that

$$E(\boldsymbol{\alpha}_{it}\boldsymbol{\alpha}'_{js}) = \Delta_{ii} \text{ if } i = j \text{ and } 0 \text{ otherwise} \quad (10)$$

where $\boldsymbol{\alpha}_{it}$ is independent of both $\boldsymbol{\delta}_i$ and $\mathbf{X}_{it}$. The asymptotic properties of their estimators are developed more fully in [72]. Even though the applications of their method [72, 73] yield plausible results, there still is lacking any evidence as to the small sample properties of the estimators described.

The only published examinations of the small sample properties of the types of estimators developed by Swamy have been by Johnson and Lyon [31] and Johnson [32]. They look at the model (1) under Swamy's simplest assumptions (2a)–(2d) and conclude that the estimator does not perform well in small samples when compared with other methods of treating models with panel data (for example, analysis of covariance). Unfortunately, their results are quite unsatisfactory in that the models that they use imply an autoregressive structure whereas the simply Swamy procedure assumes no autocorrelation. A more appropriate approach for the model they suggest would be to use a Swamy procedure assuming a covariance structure including (6) or (7) instead of $\sigma_{ii}I$ as in (4).

In a different vein, one particular problem with Swamy's estimators is that the estimators of Δ that he suggests may, at times, yield negative estimates of variances. This is so because Swamy suggests an estimator for Δ that is the difference of two matrices. Schmalensee [61] discusses this problem in detail for the simple Swamy specification but the problem still persists in the more complex situations described above. A Monte Carlo study may shed light on the possible causes of this difficulty.

HILDRETH–HOUCK-TYPE ESTIMATORS

The majority of work concerned with random coefficient estimators has been done on the so-called Hildreth–Houck model which is of the form

$$Y_t = \sum_{k=1}^{K} Z_{tk}(\beta_k + v_{tk}) \quad (t = 1, \ldots, T). \quad (11)$$

where

$$E(v_{tk}) = 0,$$
$$E(v_{tk}^2) = \alpha_k,$$
$$Ev_{tk}(v_{sj}) = 0 \; t \neq s \text{ and } k \neq j.$$

Essentially, the same model was considered by Rubin and by Theil and Mennes but H–H were the first to propose some useful general estimation procedures for the general model (11). Note that panel data is no longer necessary as was the case with Swamy-type models. The subscript t usually refers to time periods, but in appropriate situations, the H–H model may also be able to describe a single cross-sectional relationship, a situation which may be more useful in a managerial context.

In the H–H model, β_k is interpreted as the mean response of the dependent variable to a change in the kth independent variable and $\beta_k + v_{tk}$ is the actual response coefficient for observation t. In a rather interesting note, Griffiths [25] has shown that the minimum variance, linear, unbiased estimator of these actual response coefficients is not the estimator of their mean, β_k, if one uses estimation methods suggested by H–H.

Estimation of the mean response coefficients in (11) can be accomplished using ordinary least squares and this estimator is unbiased and consistent but is not efficient since the covariance matrix in the model is not scalar, i.e. $\sigma^2 I$. To improve efficiency, a generalized least squares estimator can be used, but since the covariance matrix is usually not known, one must find an estimator for this matrix. The problem thus boils down to finding the best (in some sense) estimator for the covariance matrix.

Froehlich [20, 21] and Raj [44, 45] in their doctoral dissertation and subsequent papers examine the problem of estimating the covariance matrix in the H–H model, particularly concentrating on the small sample properties of the estimators proposed.[2] A brief review of these estimators is in order.

[2] Griffiths [23, 24] also examines the small sample properties of some estimators for a slightly different H–H type model with much the same results (qualitatively) as Froehlich and Raj.

Vol. 6
1978

Firstly, it is easy to show that the covariance matrix associated with the H–H model takes the form

$$H = \begin{bmatrix} \Sigma_k Z_{1k}^2 \alpha_k & & 0 \ldots\ldots 0 \\ 0 & \Sigma_k Z_{2k}^2 \alpha_k & \cdot \quad\quad \cdot \\ \cdot & & \cdot \quad\quad \cdot \\ \cdot & & \cdot \quad\quad \cdot \\ \cdot & & \cdot \quad\quad \cdot \\ 0 & & \Sigma_k Z_{Tk}^2 \alpha_k \end{bmatrix} \tag{12}$$

The basic H–H method for estimating the α_k in H is to apply least squares to (11) and then to regress the resulting residuals on $[I - Z(Z'Z)^{-1} Z']^* Z^* = M^* Z^* = W$ where Z is a $T \times K$ matrix containing the observations Z_{tk} and where * denotes that each element of the matrix is squared (see [26, pp. 586–87] for justification of this procedure). The estimator is then

$$\hat{\boldsymbol{\alpha}} = (W'W)^{-1} W'\mathbf{r}^* \tag{13}$$

where $\mathbf{r}^*$ is the vector of squared least squares residuals from (11). H–H show that $\hat{\boldsymbol{\alpha}}$ is unbiased and consistent. However, several alternative H–H type estimators for $\boldsymbol{\alpha}$ are available. For example, H–H show that:

$$\tilde{\boldsymbol{\alpha}} = \max(\mathbf{0}, \hat{\boldsymbol{\alpha}}) \tag{14}$$

where the max is applied element by element, is biased but has lower mean squared error (MSE) than $\hat{\boldsymbol{\alpha}}$. They also propose a restricted quadratic (least squares) estimator for $\boldsymbol{\alpha}$, suggesting that one minimize the quantity.

$$(\mathbf{r}^* - W\boldsymbol{\alpha})'(\mathbf{r}^* - W\boldsymbol{\alpha}) \tag{15}$$

over $\boldsymbol{\alpha}$ subject to $\boldsymbol{\alpha} \geq \mathbf{0}$. Other estimators considered by H–H are similar to those above. For example, (13) can be replaced by

$$\bar{\boldsymbol{\alpha}} = (Z^{*\prime} M^* Z^*)^{-1} Z^{*\prime} \mathbf{r}^* \tag{16}$$

and estimators of the form (14) and (15) developed accordingly. The estimator (13) is obtained by applying least squares to the relation

$$\mathbf{r}^* = W\boldsymbol{\alpha} + \boldsymbol{\mu} \tag{17}$$

where $E(\boldsymbol{\mu}) = 0$, $E(\boldsymbol{\mu}\boldsymbol{\mu}') = \Omega$. An improvement over the estimator (13) might be possible if an estimate of Ω, say $\hat{\Omega}$, may be found and used in a generalized squares estimator of the form

$$\hat{\boldsymbol{\alpha}} = (W'\hat{\Omega}^{-1}W)^{-1} W'\hat{\Omega}^{-1}\mathbf{r}^*. \tag{18}$$

One natural estimator for Ω is discussed in [45, p. 128], namely:

$$\hat{\Omega} = 2(M\hat{H}M)^* \tag{19}$$

where $\hat{H}$ denotes (12) with estimated α's [for example, from (13)] and * again denotes that all elements are squared (see: [64] for a detailed derivation of the result $E(\boldsymbol{\mu}\boldsymbol{\mu}') = 2(MHM)^*$ using a slightly different notation).

Theil and Mennes [74] argue that the off-diagonal elements in Ω are of a low order of magnitude and so suggest replacing them with zeros and proceeding with an estimator of the form (18) which is essentially weighted least squares.

Frohlich [21] suggests using three further estimators $\boldsymbol{\alpha}$. The first is a two-step procedure, estimating Ω from restricted least squares, and then using this estimate with diagonal elements ignored to minimize

$$(\mathbf{r}^* - W\boldsymbol{\alpha})'\bar{\Omega}^{-1}(\mathbf{r}^* - W\boldsymbol{\alpha}) \tag{20}$$

over $\boldsymbol{\alpha}$ subject to $\boldsymbol{\alpha} \geq 0$, where $\bar{\Omega}$ is the above mentioned estimate of Ω from restricted least squares.

The second is a sort of MINQUE estimator (see. [48] for a discussion of this estimator) for $\boldsymbol{\alpha}$, the estimating equation for which is identical to (16) developed by H–H in another context. The third is to use direct maximum likelihood estimation as proposed originally by Rubin [58].

Raj [45] has also suggested the possible use of iterative estimators as an approach to $\boldsymbol{\alpha}$ estimation, obtaining an initial estimate of $\boldsymbol{\alpha}$, using generalized least squares to estimate the vector of mean response coefficients $\boldsymbol{\beta}$, then finding new estimates of $\boldsymbol{\alpha}$ from the generalized least squares residuals, continuing until convergence[3] is reached. An important part of Raj's

[3] Raj actually concentrates on convergence with respect to the parameter vector $\boldsymbol{\beta}$.

work revolves around the convergence properties of these iterative estimators.

No matter which estimator of $\boldsymbol{\alpha}$ one chooses, the generalized least squares estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (Z'\hat{H}^{-1}Z)^{-1}Z'\hat{H}^{-1}\mathbf{y} \tag{21}$$

where $\boldsymbol{\beta}$ is the vector of estimated mean response coefficients and $\mathbf{y}$ is the $\bar{T} \times 1$ vector of observations on the dependent variable. The Griffiths [25] estimator for the actual response coefficients is then

$$\begin{aligned}\hat{\mathbf{b}} = {} & L[Z'\hat{H}^{-1}Z)^{-1}Z'\hat{H}^{-1}\mathbf{y} + AX'\hat{H}^{-1}\mathbf{y} \\ & - AX'\hat{H}^{-1}\,Z(Z'\hat{H}^{-1}Z)^{-1}Z'\hat{H}^{-1}\mathbf{y}\end{aligned} \tag{22}$$

where L is a $TK \times K$ matrix such that

$$L = \begin{bmatrix} 1\ 1\ \ldots\ 1 & & & \\ & 1\ \ldots\ 1 & & \\ & & \ddots & \\ & & & 1\ 1\ \ldots\ 1 \end{bmatrix}. \tag{23}$$

A = diagonal $(\alpha_1, \alpha_2, \ldots, \alpha_K) \otimes I_T$ is a $TK \times TK$ matrix, and $X = (Z_1, Z_2, \ldots, Z_K)$ is a $T \times TK$ matrix with Z_k = diagonal $(Z_{1k}, Z_{2k}, \ldots, Z_{Tk})$. The small sample properties of some of the estimators of $\boldsymbol{\beta}$ have been investigated by Raj, whereas no studies are yet to be found on the small sample properties of $\hat{\mathbf{b}}$ from (22). The majority of the work of Froehlich has focused, with good reason, on the estimators of $\boldsymbol{\alpha}$, while Raj specifically looks at estimators of $\boldsymbol{\beta}$. We now turn to a discussion of their results.

FROEHLICH

Froehlich [20, 21] examines a model of the form (11) with $K = 3$, specifying that $Z_1 = 1$ for all t so that he has two independent variables and an intercept. He specifies ten separate structures (five patterns for the independent Z's, from random to harmonic, with samples of 25 or 75 for a total of ten structures), and estimates under each structure 100 times using the six estimators that we have written as equations (13), (14), (15), (16, interpreted as MINQUE), (18, as Theil-Mennes) and (20, with diagonal elements of $\bar{\Omega}$ ignored à la Theil-Mennes). For the five structures of sample size 75, Froehlich also estimates $\boldsymbol{\alpha}$ by finding iterative solutions to the likelihood equations suggested by Rubin.

In addition, Froehlich reports that several other estimators were included in his original experiment but apparently were not worth reporting in his 1973 article.

One of the tentative conclusions that may be drawn from the Froehlich experiment is that (20) is the most efficient (in terms of relative MSE) of the estimators considered. However, as Froehlich points out, this conclusion may depend upon the choice of true values for the α's and on the criterion upon which the conclusion is made. Furthermore, one might realistically ask for example whether the number of independent variables in the equation has any effect on the ranking of estimations, or, of possibly greater importance, whether the most efficient estimator for $\boldsymbol{\alpha}$ will lead to the best estimates of $\boldsymbol{\beta}$ or $\mathbf{b}$ from (21) or (22). Froehlich leaves these questions (and several others) unanswered but his work was a good initial effort at examination of a difficult problem.

RAJ

Raj [44, 45] studies two models of the form (11) but reports the results of only one of them in [45], stating that the qualitative results for the two models were quite similar. Both his models have $K = 2$, but one has no intercept term and two independent variables while the other has an intercept and one independent variable. He reports for the no intercept model.

He uses samples of either 10, 20 or 50 observations from a typical time series (so that possibly multicollinearity is a problem, although he does not examine its effect on the models used—see: [70] for a discussion in the context of Swamy-type models), two sets of parameters, and performs 100 replications for each estimator in each situation (except, without explanation, for the second set of parameters with a sample size of 50).

His main concern is the estimators of $\boldsymbol{\beta}$ as opposed to Froehlich's focus on $\boldsymbol{\alpha}$. He uses ordinary least squares (OLS) along with various GLS estimators (21) of $\boldsymbol{\beta}$ incorporating estimates of $\boldsymbol{\alpha}$ from the basic H–H method, (13)

or from (18), either with the off-diagonal elements of $\hat{\Omega}$ all zero (which he calls Theil's Weighted Least Squares—TWLS) or with the fully estimated Ω (Stepwise Aitken Least Squares—SALS). His basic conclusion is that OLS is preferable with a small sample size (10 or 20) but that SALS tends to be preferable (in terms of % root mean square error—RMSE) at a sample size of 50. In some cases he finds a slightly different ranking of estimators at alternative points in the parameter space, suggesting to us that much more work in this area may be warranted.

Raj also examines the forecasting ability of the four estimators that he examines as well as estimators for the variances of the estimated β. His results again suggest that OLS is preferable at $T = 10$ or 20 but that SALS is better at $T = 50$.

As far as the suggested iterative estimators for the model are concerned, Raj concludes that there is probably a gain in iterating the SALS, TWLS and HH estimators for a sample size of 20 but that with a sample of 50, there tends to be little or no gain in efficiency. He also points out that, at least for his experiments, the iterative estimators tend to converge reasonably well, reporting an 80–95% convergence rate in six iterations.

SUMMARY AND EXTENSIONS OF H–H MODELS

The conclusions that one might draw from the results of Froehlich and Raj suggest that in estimating the parameters of a model of the form (11), if a moderate sample size is available (over 20, say), there is a definite advantage in using a more complex estimation procedure (GLS) to estimate the mean response coefficients and to incorporate at least some information about Ω in the estimation of α to be used in $\hat{H}$ of the GSL estimator (21). Clearly much more work in this area is needed. Specifically, the effect of more than two independent variables in (11) should be investigated, a comparison of the efficiency in estimating α to the efficiency in estimating β is needed, and a look should be taken of more points in the parameter space, especially in light of Raj's result.

Recently, Singh and Ullah [65] have extended the use of the basic H–H procedure to the analysis of a set of seemingly unrelated regressions (SUR) first introduced by Zellner [80]. They conclude that their proposed random coefficient estimator (SURRC) is asymptotically more efficient than Zellner's SUR estimator if the coefficients are in fact random. Surely, an analysis of the small properties of their SURRC estimator is warranted.

The importance of further knowledge about the finite sample properties of estimators in H–H type models cannot be understated, especially since empirical studies are beginning to appear using the basic H–H framework. Singh *et al.* [64] examine the measurement of structural change in the consumption functions in seven countries while Rappoport and Kniesner [49] look at the Phillips Curve relationship in a random parameter context. It is certainly only a matter of time until other applications of the H–H model surface, since the assumption of constant coefficients in many situations has been made out of necessity and not from a belief that it was correct.

OTHER RANDOM PARAMETER SPECIFICATIONS

Our focus is on models where all parameters (or at least some of the slope coefficients) are treated as random variables. If one assumes that only the intercept is random, the resultant model is a so-called error components model, a summary of which is a paper in itself (see, for example, [2] for a bibliography and survey of some of the major results). If the parameter vector is assumed to vary stochastically between two fixed value vectors, the model is a switching parameter model, for which some estimation methods have been developed (see, for example [22]). The richness of statistical literature in the sixties and seventies on these and other specialized non-constant parameter regimes clearly points out the recognition among analysts of the narrowness of the fixed parameter assumption in applied situations.

As may be seen from the brief description in previous sections, Swamy and H–H-type models have possibly a great degree of applicability in actual applied situations. However, there are a number of other approaches to the random parameter problem that have been developed, although these are not altogether independent of the models of Swamy and H–H.

The names of Rosenberg, and Cooley and Prescott, are usually associated with these approaches, which will be outlined below.

One important contribution of Rosenberg [50, 54] is the development of what he calls the "convergent parameter" model. He assumes that in a cross section of time series the parameters of the model vary both over time and across individual units, not unlike the models of Swamy and Mehta and Hsiao, but adds the further assumption that each individual parameter vector tends to move towards a population norm parameter vector over time, hence the term 'convergent parameter'. Without going into any of the details of the model [54, 55, 56], Rosenberg has derived maximum likelihood and Bayesian estimators for the parameters of the model, and since these methods require a great deal of computation, approximate formulae based on these estimators.

A summary of the extensive Monte Carlo experiments conducted by Rosenberg with the convergent parameter model may be found in [54] with detailed results available in [55]. We will simply state some of his conclusions, referring the interested reader to Rosenberg's original work not because the model does not deserve more than a cursory discussion, but because of the detailed nature of the model and his experiments.

Rosenberg finds in the case when parameter dispersion of the type he hypothesizes is actually present, that the approximation formulae for his actual estimators work quite well except when the independent variables are serially correlated (MSE is understated), and is sharply superior to OLS especially with moderate cross section sample sizes ($N = 40$). One might conclude, then, that the convergent parameter model could be quite useful in the kinds of time series of cross section situations found in economics where the researcher suspects that the individual cross sectional units tend to conform over time to the population as a whole. What is needed at this point is a comparison of this model with the corresponding Swamy-type models which do not hypothesize the kind of convergence assumed by Rosenberg. It may be asked whether the Rosenberg methods estimate more efficiently that Swamy-type methods in situations where the convergence assumption is, in fact, incorrect. One thing that could help answer this question is a test of the convergence property, something not available at present.

Rosenberg's work has not exclusively concerned itself with the convergent parameter framework. He has examined a general Swamy-type model [52], and has applied the model to stock market data [57] and to response to drug dosage [63]. However, his major contribution seems to be the convergent parameter model.

Cooley and Prescott [12, 15] examine a slightly different kind of parameter variation in the H–H framework. They look at time series regressions and assume that parameters are adaptive in nature such that

$$\begin{aligned} \beta_t &= \beta_t^* + \mu_t, \\ \beta_t^* &= \beta_{t-1}^* + \nu_t \end{aligned} \tag{24}$$

where t refers to a time period. The basic idea is that parameters, β_t, vary randomly in any time period around some permanent parameter vector, β_t^*, and that these permanent components also vary over time. Their most recent work is a generalization of the adaptive regression model developed earlier [13, 14] which assumes only a stochastically varying intercept. One practical problem with their method is that the covariance structure for the μ_t and ν_t are assumed known, although they state that their estimation procedures show little loss in efficiency in experiments with misspecified errors [10].

As in our discussion of Rosenberg's work, we will only state a few of Cooley and Prescott's major conclusions, referring the interested reader to the original papers for detail on the exact specification of their model, estimation methods, and results of experiments. They have basically derived a maximum likelihood estimator for their model and have proved that it is asymptotically efficient [15].

Results on the small sample properties of their estimator are not yet available but Cooley [11, p. 374] reports that a large scale Monte Carlo experiment is being conducted at the NBER Computer Research Center for Economics and Management Science, and that a monograph on these experiments should be available soon. However, applications of the method with the Fair model of the US economy [11] and with a model of agricultural supply functions for the latter half of the 19th century in America [16] are quite encouraging

when compared with constant parameter specifications of the same model.

The work described above does not exhaust all of the parameter variation schemes that may be of interest to econometricians. Burnett and Guthrie [9] and Rosenberg [51] discuss a model where the parameters vary according to a stationary stochastic process with known covariance structure, although Rosenberg points out that computation requirements may be prohibitive for the model.

Sarris [59], Cooper [17] and Belsley [6] have examined to so-called Kalman Filter model from an econometric point of view and suggest estimation methods and possible applications. The basic idea of the Kalman model, first introduced by Kalman [34, 35] in an engineering context, is that the parameter vector over time obeys a first-order transition equation such that $\beta_t = A\beta_{t-1} + \mu_t$ where A is a matrix of constants.

Much of the other work on parameter variation can be considered as being closely related to the work surveyed above. Kelejian [37] has looked at the problem of random coefficients in a simultaneous equation model, while Zarley [79] at random parameters in an error in variables model. Porter [43] examines the use of random parameters in survey sampling while Belsley [4, 5] examines and suggests a test for systematic parameter variation in regression models. There are probably many others [e.g. 1, 7, 8, 18, 33, 36, 39, 40, 41, 62, 76, 78] that have not been mentioned and we should certainly see a great deal of other published work, both theoretical and applied, in the near future. In no way do we suggest that this review is complete (see: Rosenberg [53] for a survey of stochastic parameter regression—part of the special issue of the *Annals of Economic and Social Measurement*, October 1973, produced as a summary of a NBER symposium on time varying parameters).

SUMMARY

This brief survey examines the various random parameter models that have been developed in the literature (e.g. Swamy; H–H; Rosenberg; Cooley and Prescott) with particular focus on the small sample properties of the estimators developed. In general, one may conclude that the use of OLS as an estimator is relatively efficient when sample sizes are very small but that more complex estimators are superior in moderate to large sample sizes. This, of course, does not hold true exactly for all the models discussed. Of more importance is the fact that, except for possibly the H–H model, not much is known about the small sample properties of any of the estimators in random parameter models, and therefore, one should reserve the decision as to which estimator or even which model is 'best' for a particular problem until some more light is shed on these properties.

REFERENCES

1. Akkina KR (1974) Application of random coefficient regression models to the aggregation problem. *Econometrica* **41**, 369–375.
2. Arora SA (1973) Error components regression models and their applications. *Ann. Econ. Soc. Meas.* **2**, 451–461.
3. Barry CB (1977) Stochastic parameter variation in economic analysis. *J. Bus. Econ.* **29**, 139–144.
4. Belsley DA (1973) On the determination of systematic parameter variation in the linear regression model. *Ann. Econ. Soc. Meas.* **4**, 487–494.
5. Belsley DA (1973) A test for systematic variation in regression coefficients. *Ann. Econ. Soc. Meas.* **2**, 495–499.
6. Belsley DA (1973) The applicability of the Kalman filter in the determination of systematic parameter variation. *Ann. Econ. Soc. Meas.* **2**, 531–533.
7. Bowman HW & La Porte AM (1972) Stochastic optimization in recursive equation systems with random parameters. *Ann. Econ. Soc. Meas.* **1**, 419–426.
8. Brown RL, Durbin J & Evans JM (1975) Techniques for testing the constancy of regression relationships over time (with discussion). *Jl. R. Statist. Soc., Series B* **37**, 149–192.
9. Burnett JD & Guthrie D (1970) Estimation of stationary stochastic regression parameters. *J. Am. Statist. Assoc.* **65**, 1547–1553.
10. Cooley TF (1971) *Estimation in the presence of sequential parameter variation.* Unpublished Ph.D. Thesis, University of Pennsylvania.
11. Cooley TF (1975) A comparison of robust and varying parameter estimates of a macro-econometric model. *Ann. Econ. Soc. Meas.* **4**, 373–388.
12. Cooley TF & Prescott E (1973) Varying parameter regression: a theory and some applications. *Ann. Econ. Soc. Meas.* **2**, 463–473.
13. Cooley TF & Prescott E (1973) An adaptive regression model. *Int. Econ. Rev.* **14**, 364–371.
14. Cooley TF & Prescott E (1973) Tests of an adaptive regression model. *Rev. Econ. Stat.* **55**, 248–256.
15. Cooley T F & Prescott E (1976) Estimation in the presence of stochastic parameter variation. *Econometrica* **44**, 167–184.
16. Cooley TF & De Canio S (1977) Rational expectations in American agriculture, 1866–1914. *Rev. Econ. Stat.* **59**, 9–17.
17. Cooper JP (1973) Time-varying regression coefficients: a mixed estimation approach and operational limitations of the general Markov structure. *Ann. Econ. Soc. Meas.* **2**, 525–530.

18. FEIGE EL & SWAMY PAVB (1974) A random coefficient model of the demand for liquid assets. *J. Money Credit Bank.* **6**, 241–252.
19. FISK PR (1967) Models of the second kind in regression analysis. *Jl R. Statist. Soc., Series B* **28**, 266–281.
20. FROEHLICH BR (1971) *Estimation of a random coefficient regression model.* Unpublished Ph.D. Thesis, University of Minnesota.
21. FROEHLICH BR (1973) Some estimators for a random coefficient regression model. *J. Am. Statist. Assoc.* **68**, 329–335.
22. GOLDFELD SM & QUANDT RE (1973) The estimation of structural shifts by switching regressions. *Ann. Econ. Soc. Meas.* **2**, 475–485.
23. GRIFFITHS WE (1971) *Estimation of regression coefficients which change over time.* Unpublished Ph.D. Thesis, University of Illinois.
24. GRIFFITHS WE (1971) Generalized least squares with an estimated covariance matrix—a sampling experiment. Department of Agricultural and Applied Economics Staff Paper P 71–29, University of Minnesota.
25. GRIFFITHS WE (1972) Estimation of actual response coefficients in the Hildreth–Houck random coefficient model. *J. Am. Statist. Assoc.* **67**, 633–635.
26. HILDRETH C & HOUCK JP (1968) Some estimators for a linear model with random coefficients. *J. Am. Statist. Assoc.* **63**, 584–595.
27. HSIAO C (1972) *The combined use of cross-section and time-series data in econometric analysis.* Unpublished Ph.D. Thesis, Stanford University.
28. HSIAO C (1974) Statistical inference for a model with both cross-sectional and time effects. *Int. Econ. Rev.* **15**, 12–30.
29. HSIAO C (1975) Some estimation methods for a random coefficient model. *Econometrica* **43**, 305–325.
30. HURWICZ L (1950) Systems with nonadditive disturbances. in *Statistical Inference in Dynamic Economic Models* (Ed. KOOPMANS TC) 410–418. John Wiley, New York.
31. JOHNSON KH & LYON HL (1973) Experimental evidence on combining cross-section and time series information. *Rev. Econ. Stat.* **55**, 465–474.
32. JOHNSON KH (1974) On estimating models with random coefficients. *Am. Statist. Assoc. Proc. Bus. Econ. Stat. Sect.* 420–425.
33. JOHNSON SR & RAUSSER GC (1975) An estimating method for models with stochastic, time-varying parameters. *Am. Statist. Assoc. Proc. Bus. Econ. Stat. Sect.* 356–361.
34. KALMAN RE (1960) A new approach to linear filtering and prediction problems. *J. Bas. Engng* **82**, 35–45.
35. KALMAN RE & BUCY RS (1961) New results in linear filtering and prediction theory. *J. Bas. Engng* **83**, 95–108.
36. KAU JB & LEE CF (1977) A random coefficient model to estimate a stochastic density gradient. *Reg. Sci. Urb. Econ.* **7**, 169–177.
37. KELEJIAN HH (1974) Random parameters in a simultaneous equation framework: identification and estimation. *Econometrica* **42**, 517–527.
38. KLEIN LR (1953) *A Textbook of Econometrics*, pp. 216–218. Row Peterson, New York.
39. LAMOTTE LR & MCWHORTER A (1976) A test for the presence of random coefficients in a linear regression model. *A. Statist. Assoc. Proc. Bus. Econ. Stat. Sect.* 400–405.
40. MCWHORTER A, SPIVEY WA & WROBLESKI WJ (1974) Varying parameter econometric models for interrelated time series. *Am. Statist. Assoc., Proc. Bus. Econ. Stat. Sect.* 480–485.
41. MCWHORTER A, SPIVEY WA & WROBLESKI WJ (1976) A sensitivity analysis of varying parameter econometric models. *Int. Statist. Rev.* **44**, 265–282.
42. NERLOVE M (1965) *Estimation and identification of Cobb-Douglas production functions.* Rand McNally, London.
43. PORTER RD (1973) On the use of survey sample weights in the linear model. *Ann. Econ. Soc. Meas.* **2**, 141–158.
44. RAJ B (1973) *The finite sampling properties of several estimators of the linear regression model with random coefficients.* Unpublished Ph.D. Thesis, University of Western Ontario.
45. RAJ B (1975) Linear regression with random coefficients: the finite sample and convergence properties. *J. Am. Statist. Assoc.* **70**, 127–137.
46. RAO CR (1965) *Linear statistical inference and its applications*, pp. 192–193. John Wiley, New York.
47. RAO CR (1965) The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–458.
48. RAO CR (1970) Estimation of heteroscedastic variances in linear models. *J. Am. Statist. Assoc.* **65**, 161–172.
49. RAPPOPORT PN & KNIESNER TJ (1974) The illusion of the shifting Phillips curve: a model with random coefficients. *Am. Statist. Assoc. Proc. Bus. Econ. Stat. Sect.* 535–539.
50. ROSENBERG B (1968) *Varying parameter estimation.* Unpublished Ph.D. Thesis, Harvard University.
51. ROSENBERG B (1972) The estimation of stationary stochastic regression parameters reexamined. *J. Am. Statist. Assoc.* **67**, 650–654.
52. ROSENBERG B (1973) Linear regression with randomly dispersed parameters. *Biometrika* **60**, 65–72.
53. ROSENBERG B (1973) A survey of stochastic parameter regression. *Ann. Econ Soc. Meas.* **2**, 381–397.
54. ROSENBERG B (1973) The analysis of a cross section of time series by stochastically convergent parameter regression. *Ann. Econ. Soc. Meas.* **2**, 399–428.
55. ROSENBERG B (1973) Varying parameter regression in the analysis of a cross section of time series. Institute of Business and Economic Research Working Paper IP 165, University of California, Berkeley.
56. ROSENBERG B (1973) Estimation in the general linear stochastic parameter model. NBER Computer Research Center Research Report, Cambridge, MA.
57. ROSENBERG B & MCKIBBEN W (1973) The prediction of systematic and specific risk in common stocks. *J. Fin. Quant. Anal.* **8**, 317–333.
58. RUBIN H (1950) Note on random coefficients, in *Statistical Inference in Dynamic Economic Models* (Ed. KOOPMANS TC) pp. 419–421. John Wiley, New York.
59. SARRIS AH (1973) A Bayesian approach to estimation of time-varying regression coefficients. *Ann. Econ. Soc. Meas.* **2**, 501–523.
60. SCHAEFER S, BREALEY R, HODGES S & THOMAS H (1975) Alternative models of systematic risk. in *International Capital Markets.* (Eds. ELTON EJ and GRUBER MJ) pp. 150–161. North-Holland, Amsterdam.
61. SCHMALENSEE R (1972) Variance estimation in a random coefficient regression model. Department of Economics Working Paper 72-10, University of California, San Diego.
62. SCHULMAN J (1973) *A Varying Parameter Regression Study of the Trade-Off Between Inflation and Unemployment.* Unpublished Ph.D. Thesis, University of Michigan.
63. SHEINER LB, ROSENBERG B & MELMON KL (1972) Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comp. Bio. Res.* **5**, 441–459.
64. SINGH B, NAGAR AL, CHOUDHRY NK & RAJ B (1976) On the estimation of structural change: a generaliza-

tion of the random coefficients regression model. *Int. Econ. Rev.* **17**, 340–361.
65. SINGH B & ULLAH A (1974) Estimation of seemingly unrelated regressions with random coefficients. *J. Am. Statist. Assoc.* **69**, 191–195.
66. SUNDER S (1975) Stock price and risk related to changes in inventory valuation. *Acct. Rev.* **50**, 305–315.
67. SWAMY PAVB (1968) *Statistical inference in random coefficient regression models.* Unpublished Ph.D. Thesis, University of Wisconsin.
68. SWAMY PAVB (1970) Efficient inference in a random coefficient regression model. *Econometrica* **38**, 311–323.
69. SWAMY PAVB (1971) *Statistical inference in random coefficient regression models.* Springer, Berlin.
70. SWAMY PAVB (1973) Criteria, constraints and multicollinearity in random coefficient regression models. *Ann. Econ. Soc. Meas.* **2**, 429–450.
71. SWAMY PAVB (1974) Linear models with random coefficients, in *Frontiers in Econometrics* (Ed. ZAREMBKA P) pp. 143–168. Academic Press, New York.
72. SWAMY PAVB & MEHTA JS (1974) Estimation of linear models with time and cross-varying coefficients. Unpublished.
73. SWAMY PAVB & MEHTA JS (1975) Bayesian and non-Bayesian analysis of switching regression and of random coefficient regression models. *J. Am. Statist. Assoc.* **70**, 593–602.
74. THEIL H & MENNES LB (1959) Multiplicative randomness in time series regression analysis. *Econometric Institute Report* 5901, *Netherlands School of Economics.*
75. THEIL H (1968) Consistent aggregation of micromodels with random coefficients. Center for Mathematical Studies in Business and Economics Report 6816, University of Chicago.
76. THEIL H (1971) *Principles of Econometrics.* John Wiley, New York. pp. 570–573, 580–587, 622–628.
77. WALD A (1947) A note on regression analysis. *Ann. Math. Stat.* **18**, 586–589.
78. WELSCH RE & KUH E (1974) The variances of regression coefficient estimates using aggregate data. *Econometrica* **44**, 353–363.
79. ZARLEY C (1975) *Statistical Estimations in a Random Coefficient Errors in Variables Model.* Unpublished Ph.D. Thesis. University of Kansas.
80. ZELLNER A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Statist. Assoc.* **57**, 348–367.
81. ZELLNER A (1966) On the aggregation problem: a new approach to a troublesome problem. Center for Mathematical Studies in Business and Economics Report 6628, University of Chicago.

ADDRESS FOR CORRESPONDENCE: *Dr LW Johnson, Macquarie University, Dept of Economics, North Ryde, New South Wales 2113, Australia.*

1. 6
78

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 83–88, 1978. Pergamon Press. Printed in Great Britain

Planning Models for Budgeting Teaching Resources

LD SMITH

University of Missouri, St. Louis

(*Received March 1977; in revised form June 1977*)

This paper relates experience in the development of a set of integrated planning models for a school board in a public school system. Three modules were developed. An enrolment projection module utilizes a modified Markov transition matrix (augmented by adaptive exponential smoothing models to forecast flows into the system). A program review module operates in interactive mode as a deterministic simulation to generate alternative budgets in a 'planning–programming–budgeting' context. A resource planning module was produced as an optimizing goal programming formulation, but was not found useful to implement for reasons outlined in the paper.

INTRODUCTION

RECENT declines in the growth of enrolments in educational institutions have created particular problems when budgeting for new program options. In periods of growing enrolments it was possible to initiate programs and acquire the necessary teaching resources with the help of expanded budgets. As budgets now become more restrictive in response to stabilizing or declining enrolments however, there is an apparent need for a convenient method of reviewing the financial impact of possible changes in program mixes and of changes in resource allocations to specific programs in a mix. In response to this need, a set of computerized models have been developed to support the planning function of local school boards responsible for elementary and secondary school education (usually kindergarten to grade 12).

In this paper the term 'program' refers to a specific segment of a student's educational experience for which a distinctive allocation of teaching resources is required. Examples of programs would be mathematics, science, art, music, physical education, independent study, remedial reading, or guidance counselling. The term 'teaching resources' encompasses all nonadministrative teaching personnel. Examples of teaching resources would be guidance counsellors, specialists in art, mathematics, physical education, music, or industrial arts, and (in elementary grades) regular class teachers. School supplies and administrative personnel are presumed to be budgeted separately to the extent that tradeoffs among the three broader resource categories would occur only in terms of total budget allocation to each, and would not, for example, involve a choice between an art instructor and a secretary to the vice principal.

The computer models are integrated into a package consisting of three modules. An enrolment projection module is designed to enable the board to specify separate factors which influence future enrolments in the school system, and generate enrolment projections based upon alternative assumptions about the relevant factors. A program review module enables an investigation of the budgetary implications of operating programs under different resource commitments. Finally, a resource management module considers the problem of adjusting the teaching resources to levels required to support programs anticipated over ensuing years.

ENROLMENT PROJECTION MODULE

Future enrolments in a school system result from two types of flows: (1) flows within the

system consisting of students who previously have been enrolled in the board's school programs and (2) flows into the system consisting of students who have not previously been enrolled. The first year of schooling would of course consist almost entirely of flows in the latter category. Beyond the first year of schooling flows in the former category dominate and fortunately can usually be extrapolated reasonably accurately.[1]

The basic enrolment projection model

The basic model chosen for projecting enrolments can be stated as:

$$E_{jt} = F_{jt} + \sum_i P_{ij}(t)E_{i,t-1} \quad (J = 1,2,3,\ldots) \qquad (1)$$

where

E_{jt} is the number of students enrolled in grade j during period t;

F_{jt} is the net flow of students from without the school system into grade j during period t;

$P_{ij}(t)$ is the proportion of students enrolled in grade i at time $t-1$ who return to enroll in grade j at time t.[2]

The flows F_{jt} $(j = 1,2,\ldots)$ reflect immigration of student population into the school district, changes in the age distribution of migrating residents, and the propensity of new students in the community to select schools in the system. F_{1t} includes, in addition, the effect of community birth rates. The Markov transition probabilities $P_{ij}(t)$ reflect the rate of progress of students through the school grades (pass/fail rates), the tendency of students in the higher grades to discontinue formal education (drop-out rates), and the propensity of students to emigrate from the system.[3] To generate forecasts of enrolment, estimates must first be obtained for future values of F_{jt} which can then be appended to the recursion of flows within the system as described in equation (1).

Estimation procedures

Estimation procedures to be utilized in forecasting F_{jt} would depend upon the extent of disaggregation afforded by available data. Preferably one would be able to specify separately the different components affecting F_{jt}, as for example:

$$F_{jt} = f_{jt}M_{jt} + T_{jt} \text{ for } j = 2,3,\ldots \qquad (2)$$

and

$$F_{1t} = f_{1t}B_{t-6}P_{t-6} + f_{1t}M_{1t} \qquad (3)$$

where

M_{jt} is the net migration of children into the district who are eligible to enroll in grade j during period t;

f_{jt} is the fraction of students eligible for grade j in period t who select the school system over alternative school systems in the district;

T_{jt} is the transfer of students into grade j in period t from alternative school systems in the district;

$B_{t'}$ is the birth rate for residents in the district during period t';

$P_{t'}$ is the district population during period t'.

The enrolment projection equations (1)–(3) can be used to generate forecasts under different assumptions about future migration rates, birth rates, population levels, and rates of transfer among alternative school systems in the district.

To help cope with situations where data limitations preclude the use of equations (2) and (3), one program available in the forecasting module takes total enrolment data for each grade and extracts an estimate of net flow into the system using

$$F_{jt} = E_{jt} - \Sigma_i P_{ij}E_{i,t-1} \text{ for } j = 1,2,3,\ldots. \qquad (4)$$

The values used for P_{ij} are the averages (weighted by number of students if desired) of annual empirical proportions of students in grade i who returned to enroll in grade j for the following year. Resulting P_{ij} values are then used to develop probability distributions for the grade level of a student after n periods.

[1] Flows within or out of the system can even be controlled to some degree by changing criteria for advancement or by dedicating special resources to students with problems.

[2] In most cases the transition probabilities $P_{ij}(t)$ seem to remain quite stable over time enabling the time qualifier to be dropped and $P_{ij}(t)$ to be replaced with P_{ij}.

[3] Treating flows within the system as a time-dependent first order Markov chain is not actually justified on theoretical grounds because an individual student's probability of advancement seems dependent upon his educational history. Since failure rates are not usually very high however, little improvement in forecasting accuracy could be anticipated from using higher order Markov-chains which account for recent educational history. Balachandran [1] discusses some alternative forecasting models.

Vol. 6
1978

These distributions are compared with empirical distributions of total time until graduation. If discrepancies become apparent, the P_{ij} estimates are modified with consultation from the school principals, time to graduation distributions are developed again, and the cycle repeats itself until the implied distributions of total time to graduation and the individual P_{ij} values both appear reasonable.

F_{1t} is always considered separately, but experience has suggested that it is advantageous to group remaining flows as

$$F_t = \sum_{j=2}^{12} F_{jt}.$$

Then $\hat{F}_{jt} = f_j \hat{F}_t$ (where f_j is the fraction of net flow into the system which falls in grade j) can be used to estimate flows in particular grades.[4]

Another program uses exponential smoothing for subsequent extrapolation of flows. An option enables the user to select an adaptive or nonadaptive exponential smoothing model. The former type includes a 'tracking signal' to adjust the value of the smoothing constant automatically in response to the forecast errors experienced. Both models produce an estimate of the current smoothed average of flows and an estimate of trend in the smoothed average to enable extrapolations several periods ahead.

A summary of factors to consider when forecasting enrolments

Anticipation of extraordinary events such as the termination or expansion of alternative educational facilities in the district, or unprecedented commercial and residential construction in the area obviously call for adjustments to be made in any forecasts generated from historical data alone. Regardless of the specific form chosen for the forecasting model however, one should investigate possible changes in each of the following factors and consider their net effects upon future enrolments:

1. birth rates—responsive to changes in social values and the tendency for women to select working careers outside the home.

2. migration tendencies—responsive to changes in the economic and social environment.

3. age distribution of the population in the community—responsive to changes in the economic climate and to the type of residential construction to be undertaken (e.g. multiple unit dwellings tend to accomodate families with children in the elementary vs advanced grades).

4. availability and cost of alternative educational opportunities—future developments in the public and private school systems which might change their relative attractiveness.

PROGRAM REVIEW MODULE

The program review module serves as the core of the planning exercise. The following attributes are considered essential to its success as a planning tool:

1. Decision orientation—the variables under direct or indirect control at the board level are incorporated separately, enabling the analyst to change one factor (e.g. increase class size utilized in delivering a program), make simultaneous compensatory changes in another factor (e.g. teachers' salaries or allotment of teachers' aides), and quickly obtain information about the net effect of the changes upon the board's budget.

2. Communicability—information required for the specification of model parameters is objective and well defined. School board personnel can provide data readily without learning technical 'jargon'. Similarly, the model output is expressed in familiar understandable terms.

3. Interactive capability—The model is a simulation model (nonoptimizing) designed to enable the analytical talents of the decision maker and the computational talents of the computer to converge upon the planning problem simultaneously. The decision maker is able to analyze the resource implications and budget requirements which would result from operating a set of programs in a specified manner. He can then change one or more factors in order to alleviate perceived problems, obtain a new solution immediately, examine it, and continue the cycle until a satisfactory set of alternative plans is devel-

[4] $f_j = (1/11)$ has been used in a 12 grade system unless cumulative data supports the conclusion that f_j $(j = 2, 3, \ldots 12)$ is nonuniform in shape.

oped. The planning and budgeting process can thus be undertaken by a team of individuals who can respond to tentative solutions to problems, suggest alternatives, and obtain immediate feedback about the financial viability of their proposals. This approach stands in contrast with the use of a linear programming formulation advocated by Harman [7]. In the latter case, a search for an optimal solution occurs within guidelines specified in advance and formulated in terms of the LP constraints. Our experience suggests that the interactive models are much more effective in stimulating further exploration of decision alternatives, that team involvement is easier to sustain, and that planning proceeds with less rigidity than experienced when large optimizing models are used.

4. Responsiveness to specific proposals—many of the school board's decisions about program offerings, teachers' working conditions, etc. are made as a result of negotiation and compromise among groups with different special interests. Compromises are often negotiated without a complete analysis of the overall financial impact of proposals simply because there is a tight time constraint upon the budgeting process. The interactive model makes it possible to scrutinize proposals from opposing parties and determine their overall financial impact, in a manner which encourages consideration of a wide range of possible trade-offs.
5. Primary focus on programs rather than budget constraints—a set of school programs is tentatively designed to meet the board's educational objectives, the 'line budget' totals necessary to sustain the programs are then determined, and the programs are subsequently modified as necessary to produce feasible budgets for the line items. The program review module thus represents a 'planning–programming–budgeting' system in which the "programs" are defined in a workable nonoverlapping manner as suggested by Hartley [4].

DEFINITION OF VARIABLES AND PARAMETERS

Values for variables and parameters of the model are specified with the help of a questionnaire and work sheet. Definitions and explanations of model variables and parameters appear below and are followed immediately by a summary of the mathematical operations performed to generate the reports.

Let:

k be the number of different grades in the school system. Even 'nongraded' modular school programs can utilize grade definitions to stratify samples in estimating enrolments for each program.

L be the number of different types of teaching resources.

p_k be the number of different programs offered exclusively to students in grade k.

N be the number of nongraded (open to students from more than one grade) programs offered.

g_k be the number of different academic tracks defined for students in grade k. Students in the same academic track tend to choose similar course options—e.g. students in the science track would have a heavier concentration of science electives than students in the arts and languages track.

d_{ik} be the proportion of students in grade k who enroll in academic track i.

E_k be the total expected enrolment of students in grade k. These estimates are obtained from the first module in the planning system.

S_{ijk} be the number of hours spent each week in program j of grade k by students enrolled in academic track i of grade k.

r_{ekj} be the number of student-contact hours of resource e required to deliver each class hour of program j in grade k. It is possible that more than one resource person is available to a class. For example, teachers' aides may be allocated in such a way that each hour of music class receives 1 hr of time from a music specialist and one-third hour of time from a teachers' aide.

C_{kj} be the size of class utilized in delivering program j of grade k.

Z_{ink} be the average number of hours spent each week in nongraded program n by students in the i^{th} academic track of grade k.

Y_n be the size of class utilized in delivering

Vol. 6
1978

non-graded program n. Some programs such as remedial reading or counselling are handled on a one-to-one basis. In these instances Y_n is simply set equal to 1.

W_{en} be the number of student contact hours from resource e required to deliver each hour of non-graded program n.

T_e be the number of student contact hours provided each week by a full-time staff member in resource category e. Only hours which are formally allocated to graded or nongraded programs are to be included.

Q_e be the average annual salary of staff members in resource category e.

Each of the items above is designated a value by the user. The following items are calculated from the data provided.

V_{ekj} is the total number of teaching hours of resource e devoted to delivering program j in grade k each week.

$$V_{ekj} = (r_{ekj} \cdot E_k / C_{kj}) \cdot \sum_{i=1}^{g_k} d_{ik} S_{ijk}.$$

X_{en} is the total number of teaching hours of resource e devoted to delivering nongraded program n each week

$$X_{en} = \left(W_{en} \cdot \sum_{k=1}^{K} \left(E_k \cdot \sum_{i=1}^{g_k} d_{ik} Z_{ink} \right) \right) \Big/ Y_n.$$

R_e is the total number of full-time staff equivalents required of resource e.

$$R_e = \left(\sum_{k=1}^{K} \sum_{j=1}^{P_k} V_{ekj} + \sum_{n=1}^{N} X_{en} \right) \Big/ T_e.$$

B is the total budget for nonadministrative teaching personnel.

$$B = \sum_{e=1}^{L} R_e \cdot Q_e.$$

Different breakdowns of the budget can of course be obtained simply by summing relevant variables over the appropriate subscripts. Currently, summaries are produced by resource and by program. Simple modifications would enable program components to be grouped by administrator also.

Operation of the program review module

The program review model is programmed in BASIC in order to be readily adaptable to implementation with a mini-computer. Development and testing were actually performed partially upon a Hewlett-Packard 2000A computer (16K memory) and partially upon an IBM 370/165 using TSO. Preliminary applications seem to confirm our contention that school board officials feel much more comfortable using models which process information to demonstrate the resource implications of alternative program specifications, than they feel using optimizing models which demand a specification of the complete set of alternative strategies before they have an opportunity to review the impact of a number of intuitively reasonable strategies. Reviewing tentative solutions did sharpen the decision makers' appreciation of different trade-offs possible and did motivate consideration of alternatives which were not contemplated in advance. Furthermore, the program planning module can be employed without the technical background required by users of optimizing models.

RESOURCE MANAGEMENT MODULE

With the help of the program review model, a tentative set of programs and their associated resource requirements can be developed for future periods (a 2–5 year time horizon seems popular). The resource management problem involves manipulating the teaching staff in each resource category in order to provide the skills required to operate the planned programs. The board often finds itself with a surplus of skills in one category and a deficiency in another. Furthermore, normal attrition of personnel in the surplus categories is unlikely to alleviate the problem substantially, and there is a reluctance to discharge personnel (often tenured with relatively high salaries) who have served faithfully in the past, but who find their skills less in demand in the current environment.

This leaves the board to select some combination of the following options:

1. encourage personnel in the surplus categories to acquire new skills which would qualify them to offer their services in support of alternative programs.

2. overstaff programs which utilize the skills of personnel in the surplus categories and operate the new programs at reduced staff levels, or perhaps delay the introduction of new

programs until attrition of current staff enables hiring specialized personnel for the new programs.

3. discharge surplus personnel in accordance with some negotiated priority scheme.

Following Dyer [2] and Lee and Clayton [5], it was felt that the problem could conveniently be analysed within a goal programming framework. However in attempting to adapt the goal programming methodology to the affairs of a school board, the following problems were encountered:

1. The series of links between personnel shortages or surpluses, their subsequent effects upon the quality of programs under review, and the effects of program changes upon the quality of education received by the students are neither well-defined nor objectively measurable.

2. Changes contemplated in staff levels for resource categories which support new programs frequently involve the manipulation of only one or two staff, thus calling for integer restrictions upon variables which are small in magnitude.

3. The types of costs incorporated in the objective function to account for the disutility of different deviations from goals are not homogeneous and were difficult to transform into a single dimension.

4. Options which could be selected in adjusting staffing were highly dependent upon the adaptability of specific personnel, and staffing was interrelated with scheduling.

It became evident that developing a goal programming formulation which would lead to an effective solution to the problem was possible only by the incorporation of a number of additional constraints which did nothing more than reflect the logic of the decision maker as he systematically selected and eliminated alternatives. By the time sufficient information had been included to produce a reasonable goal programming formulation, the necessary strategies for addressing the real problem had been defined, and the goal programming model itself was redundant. In other words, the set of alternatives at hand and the effects of each were so complex and specific to the individuals who were available to staff the school programs, that specification of an optimizing model to enable an automatic review and evaluation of alternatives was not perceived to be worthwhile.

CONCLUSIONS

The forecasting module and program review module of the planning package were accepted readily and proved to be useful tools for school board planning. The resource management module failed to be a useful tool in its present form. Upon reflection, it appears that an attempt was made to apply a conceptually useful tool which is simple in form to a problem with many complexities which defy measurement. The resource management phase appears to require a good procedure for assimilating information when reviewing the effects of broad strategies for coping with the problem, but not an automatic method of finding the best strategy once the information is available. The resource management module is being redesigned accordingly.

ACKNOWLEDGEMENT

We wish to express our appreciation to Dr. John Moore of the Kirkwood School District, St. Louis, MO for helping test the planning models and suggesting modifications to improve the package. He was most influential in reorienting our thinking on the resource management problem.

REFERENCES

1. Balachandran KR (1973) Variable-work models for predicting course enrollments. *Ops Res.* **21**(3), 823–834.
2. Dyer JS (1973) A procedure for selecting educational goal areas for emphasis. *Ops Res.* **21**(3), 835–845.
3. Harman WG (1971) *Policy Models for Planning Teacher Manpower*, (unpublished Ph.D. thesis, University of Toronto).
4. Hartley HJ (1968) *Educational Planning–Programming–Budgeting*. Prentice-Hall, Englewood Cliffs, NJ.
5. Lee SM & Clayton ER (1972) A goal programming model for academic resource allocation. *Mgmt Sci.* **18**(8), B395–B408.

ADDRESS FOR CORRESPONDENCE: *L Douglas Smith Esq, School of Business Administration, University of Missouri, 8001 Natural Bridge Road, St Louis, MO 63121, USA.*

MEMORANDA

ol. 6
978

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 91–92, 1978. Pergamon Press. Printed in Great Britain

Methodology and Teleology

(Based on an address by the President of the Hellenic Republic of Greece to the 1977 TIMS Conference in Athens)

THE ANCIENT Greeks had realized, despite the fact that life in those days was much less complex and involved, that common sense and simple experience do not suffice when one is planning to organize thought or action; what is required is true knowledge. True knowledge can only be acquired through the rationalization of things, their organized and systematic conception. Only in this way does thought become clearer and sounder, and action more effective, swift, and economical. Only in this way can we fully exploit the material at our disposal. Nothing is wasted, nothing remains unexploited.

In this process, as men have known since those days, the greatest aid is the number. Every time we conceive and express quality as quantity, our knowledge increases, and along with it our powers of thinking and acting correctly.

'*Methodos*' is a Greek word—'*odos*' means road. Method is the road we follow. Method is a course consciously followed. For every activity a method defines the way of management. A research of management is also research of the method which defines the management.

Since everywhere and at every time, thought and action require method and management, it is natural to wonder whether there are some common methodological principles governing every kind of activity.

Let us ignore, in this series of thoughts, the methodologies of the different sciences and the principles governing them. There is one methodology for the physical/mathematical sciences with special laws ruling; a different methodology for the social sciences, another for history, another for the legal sciences. Let us pause at the infinite forms of action. There we might find some guiding pieces of advice which would be of value in all cases. For instance, we could transform, to the degree that it is possible, the given facts from qualitative to quantitative ones, weigh the endurance of the social matter we have to handle, weigh its reactions, and act with all these in mind; we should not forget that action is accomplished on material which is not immobile but under constant motion; we should not forget that time is a measurable factor of every accomplishment, and finally we should not forget that whatever is accomplished by human beings in human communities has infinite facets. If the value of the accomplishment is to be judged with any degree of accuracy, the main ones of these facets must be practicable for the person involved.

It goes without saying that towards this end, it must be understood by all, that the experiences of individuals are limited and that on each problem of management light is shed from many sources; in other words, it is more thoroughly lit when many individuals with varied knowledge and experiences collaborate, so that from the dialogue or from the conflict of many opinions, truth will shine more brightly forth.

All these are correct methodological recommendations, but they are not methodological laws. Do such laws exist? I will dare to mention one which is basic, so basic that it may be taken for granted: the law of 'the final cause'. These numerous studies of the various facets I mentioned are all very well. Their value, however, will be judged by the extent to which they serve the pursuit of each particular cause. The captain at sea studies and organizes many different data. Towards which final cause? In order to steer his ship as fast and as safely as possible to the port of his destination. This is why, while studying the infinite data necessary to the management of his journey, he will raise his eyes towards the polar star, which in our days has been replaced by electronic equipment. I am trying to say, using this image, which may not be altogether adequate, that the person who acts, must never forget the final goal of his action. That this goal must be, in every way, the last and decisive criterion of this action. We can consider this a law.

I will add, further, that when many individuals collaborate towards the solution of the same problem—which is the way it should be—dialogue can be of use only if all have agreed on the final cause, if it has become clear to all that they are pursuing the same goal. From thereon the disagreements concerning the most suitable means of pursuing the goal become truly useful in the procedure leading to the correct solution.

We, the people of today, are so overwhelmed by the countless technological, statistical and managerial means at our disposal, that we flounder, and sometimes are so bewitched by these spectacular means, that we forget our final cause. Sometimes we mistake some of these spectacular means for the goal itself, while the real one lies elsewhere. But I will go one step further. Is it, I wonder, certain every time that this is our 'final cause'? Could this too be simply a means towards an ultimate end which has been forgotten in the course of all these procedures? I am afraid that this is what usually happens. We do not exhaust the full scale of means and ends in order to reach an ultimate goal, one of which we could say that for this and this only all the actions of a society of free and civilized people, are worth their while. This we consider superfluous or take for granted, whereas it is the ultimate criterion of what must be done, and how one must do whatever it is he is doing. We do what those who roam over the countryside do: through too much preoccupation with the trees, they lose sight of the forest. They lose the image of the whole.

So, in order to sum up all this chatter with a more scientific phrase, I say that the basic law for one engaged in action is the law of the 'final cause', the law of the '*telos*', and the method he will apply in order to find his way is the '*teleological*' method. Instead of thinking causatively, as in the natural sciences, in the final phase, when, that is, composing all his data, he

will think *teleologically*. So if we wanted to lay a sound foundation for the management of any social activity we would have to base our study on *the law of the final cause*—the final cause of each one of our endeavours, and submit each separate task, each study, each action, to the criterion of the final cause, as to a touchstone.

It is hardly necessary to stress that all these guiding methodological principles cover an infinite and varied material. Whether you are going to organize a dance or a hospital, an oil refinery, an educational institution, a municipality, a State, or a confederation of States, while the content, the matter, may differ, a total of methodological procedures, which remain the same in all these cases, does not change. The matter changes but the form remains. Naturally this is true only for several general terms, such as the one mentioned above; there are some more specialized rules which cover only a more limited system of similar phenomena. However, that which mainly elevates our work to the level of sciences is the existence of several general methodological principles which cover any matter. For this reason only the system of your experiences and your thoughts has educational value and subjects all the separate experiences to several permanent and general (requests), to a systematic whole, which in itself may be the object of a science.

I am afraid that you know all this better than I. Preoccupation with a multitude of trees does not rob you of knowledge of the forest, the surveillance of the whole, and this is the reason for your personal success, in whatever field you tested your abilities.

You must excuse me if I spoke drearily, and spent too much time on what may be considered commonplace. I have spent many years as a teacher and have acquired the professional deformation. But before I finish I will tell you a few words, not as a teacher, but as a man of action, a man of action of a certain age.

I am a fervent supporter of the tendency to replace the common experience, the common sense, with a strictly organized system of thought, possibly expressed through mathematics. But life, its variety and its infinity, however much we may want it, is never conceived in its totality. Something remains inconceivable, it rests beyond the limits of our knowledge—no matter how great a progress we make in the management of our thought and our activities. Nevertheless life calls us to action, even without complete knowledge. What happens then? Then, from the depths of the consciousness of the most privileged human beings—and usually during their most difficult moments—sounds a voice indicating the road, a sense—incalculable and elusive, which guides us in the dark. It shows the right road, leading to the right goal, only to the most privileged. What of the others? If they are brave, they will take risks; if not, they will probably stop looking.

Happily your researches are reducing this vacuum, reducing the dangers, and leading each man of action with safer tread. So long as he glances at the polar star every now and then, to check his course towards his port of destination, towards his goal, which must finally serve not only himself, but broader groups of people, and ultimately humanity itself.

In all this endeavour to rationalize action, in order to manage it, so that it can be accomplished methodically, you, with your studies and research, with today's Congress, are offering a very valuable service; one which is particularly appreciated here in Greece, in the land where that which we call rational thought and scientific method were born, where myth and dogma gave way to organized scientific research and strict mathematical thought.

I am happy that TIMS chose Athens as the spot to congregate in 1977—the ground that holds the roots of the great tree which you nourish and tend throughout the civilized world. The return to the roots and the awareness of its basic principles does not mean returning back; it is, on the contrary, the best way to stride towards new conquests, with a surer tread.

His Excellency Constantine D Tsatsos
(July 1977)

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 93–96, 1978. Pergamon Press. Printed in Great Britain

The Student Nurse Allocation Problem: A Formulation

THE PROBLEM of designing effective student nurse training programmes has received a good deal of analytic attention [see 4, 7, 8, 9, 10] which has not yet, apparently, yielded an approach acceptable to those responsible for producing such programmes. The problem of determining programmes which produce a balance between service and teaching needs is extremely cumbersome to handle manually partly on account of its inherent combinational characteristics and partly because each hospital poses a unique set of conditions. The nature of the problem is to produce a 3 year time-table for student nurses (a similar problem is encountered for pupil nurses) which complies with the minimum practical and theoretical standards laid down by the General Nursing Council (GNC) while, at the same time, attempting to utilise the students as an integral part of a hospital labour force.[1]

Inevitably, these dual requirements can be in conflict for at least part, if not all, of the duration of the trainee's course, the conflict arising from the fact that the basic academic training takes place away from the ward in a formal teaching process. A student's time is consequently divided into three main sections: school block periods, holidays and ward allocations. An example of a student nurse programme for a hospital with two intakes is set out in Fig. 1 below.

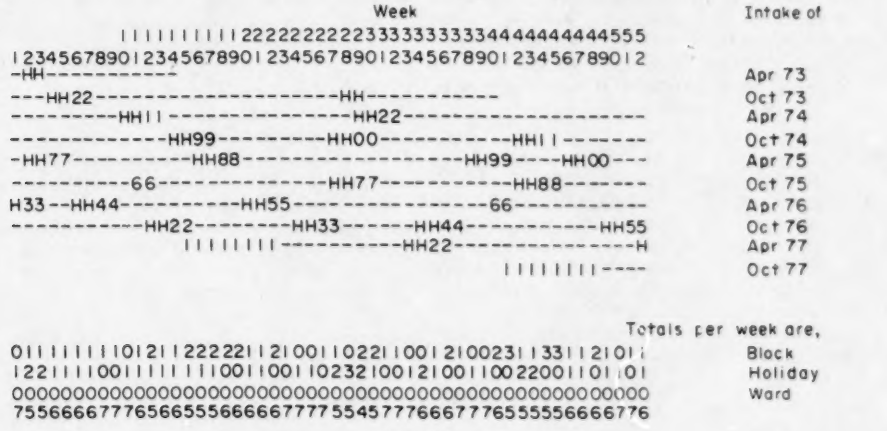


FIG. 1.

The particular course illustrated lasts for 3½ years and enables an entrant to secure both a general (SRN) and a children's (RSCN) nursing qualification. Within this period a student undergoes 30 weeks of in-school training in twelve separate sessions. She (or he) is entitled to 22 weeks holiday during the period of the course while, for the remainder of the time, the student is available for ward allocation. In Fig. 1 holiday periods are represented by 'H', ward allocation by '–' and in-school periods by the number of the block being undertaken.[2] The arrival month and year of each intake represented is listed on the right-hand side. The totals along the bottom of the figure show the total number of intakes in block, on holiday, or on ward allocation for each week of the year. Thus the October 1977 intake would commence their training in week 41 and the first eight weeks of their course would be spent in the school of nursing. In the 12th and 13th weeks of 1978 they would be on holiday which would be followed by a fortnight in school. Eight weeks of ward duty would be followed by another 2 weeks holiday and 2 weeks school, etc.

Previous analytic work in this field has been concerned either with the formulation of specific rules to aid the manual allocators, or with examining the scope for the use of a computer in the production of student nurse allocation programmes (SNAP's). The most comprehensive work in the first area is undoubtedly that undertaken by Banks [1], his prime conclusion being that the design of the block programme itself is the major factor affecting a successful allocation. With this in mind he then produced block programmes which were primarily aimed at ensuring that when one batch of trainees left a ward they were replaced by a batch immediately their junior. This, of course, has the attraction of providing some measure of stability in the level of experience of trainee staff on a ward.

This concept has much inherent appeal and works well in situations where its implementation is feasible. If a hospital has three or four intakes per year then allocation periods of 13–18 weeks are possible. These not only accord with GNC requirements but they also

[1] The latest manpower returns reveal that 29% of the nursing labour force is trainees and it is common to find in excess of 60% of the bedside care provided by such unqualified staff.

[2] When block numbers exceed nine they are represented by their final digit only.

lend themselves to 'cross-over' allocation.[3] However, in these situations where there are only two intakes of students per year securing this desirable 'cross-over' objective would require allocation periods of approx. 6 months which are infeasible due to the 3 year duration of the course. At the other end of the scale five intakes per year would result in ward allocations of less than 8 weeks which would violate GNC requirements.

The computer approach to student nurse allocation was initiated in the mid 1960s. Typical of this early work [see 2, 5, 6] is that carried out by the Operational Research Unit of the then Oxford Regional Hospital Board [3]. Details relating to each student nurse and all training locations in addition to the number of weeks experience required in each speciality were stored. The computer was programmed to select each nurse who was scheduled to complete the required number of weeks in the speciality of her present assignment. Details of her past experience including her availability and non-availability for each speciality were then printed out (if she had completed the minimum number of weeks on a speciality she became 'not available' for that speciality). When 'approved' by a nurse administrator the list (amended as required) was fed into a second programme which allocated jobs to nurses in an available speciality. A new list showing the nurses to be changed, along with their new job, was produced. This list was, to all intents and purposes, equivalent to the 'change-list' which traditionally informed trainee nurses of their moves. One of the major limitations of this approach was its myopic characteristic which was very likely to produce backlogs and bottlenecks of nurses needing experience, possibly all in the same speciality, particularly near the end of the 3 year training programme. With the apparent failure of these 'clerical computer' systems something of a 'backlash' against the use of computers in the resolution of this problem undoubtedly occurred. There seemed to be a general feeling that the complexity of the problem ruled out the possibility of using a mechanistically based approach. However, it is the authors' belief that a problem of this nature can and does lend itself to the rapid computational facility provided by a computer.

The student nurse allocation problem consists of two sequential parts namely the production of a block programme and the allocation to particular specialties and it can fairly readily be formulated as a sequencing problem.[4] A series of in-school periods (blocks) and holiday periods must be time-tabled into 3 years according to rules laid down by the GNC. Within those stipulated rules there exists some flexibility within which each individual allocator may operate. This flexibility can be translated into *earliest* and *latest* start times for each block period. If variables are then assigned to these possible start dates and constrained to take on the value of 0 if the block does not start at that particular week, and 1, if it does, then the design of the block programme becomes a mathematical *programming* problem of the zero–one variety. Before considering the manner in which the multi-faceted aspect of the problem can be dealt with, it is useful to outline how the linear constraint set was formulated. This constraint set defines the 'limits' within which any solution must lie. Given that any particular allocator defines his or her own requirements according to local circumstances, and their own personal views on programme design, that detailed below must be seen as being purely illustrative.

Firstly fifty-two constraints are set up (one for each week of the year) which guarantee that the number of student intakes in block at any one time does not exceed the classroom or teaching capacity of the particular school involved. Another fifty-two constraints are necessary to ensure that the number of student nurses available for ward allocation is as constant as possible. It is interesting to note that if one is dealing with equal numbers and constant wastage rates for each intake and all of the fifty-two nurse availability constraints are satisfied exactly (i.e. an equal number of nurses are available for ward allocation each week) then one produces the same programme as Banks's manual rules perfectly applied would have done. However, as noted previously, the conditions necessary for such an occurrence make his technique a limited special case of this more general approach.

The final set of constraints specify the minimum allocation period between blocks to provide adequate time for practical experience as set in GNC requirements. It is clear that if a constraint is required, which minimises fluctuations in the number of nurses available for ward allocation, this is directly equivalent to minimising fluctuations in the total number of nurses on holiday and in block during each week. Thus the constraints must be formulated so that over a one year period a 'preferred' number of nurses are either in block or on holiday each week.

The value of the 'preferred' number of nurses NOT available for ward allocation is calculated as follows. During any year each intake will be represented at some time during that year by a number of identical intakes which have arrived during both past years and the current year and each of which is at different stages of their training. The total number of weeks of experience achieved by these portions of an intake during the year will be equivalent to one full term of experience of that intake.

Hence:

No. of nurse block weeks occupied per year
$$= \epsilon_i \epsilon_j [N_i * K_{ij}] = P \text{ (say)}.$$

No. of nurse holiday weeks occupied per year
$$= \epsilon_i \epsilon_j [N_i * L_{ij}] = Q \text{ (say)}.$$

Where:

N_i = No. of nurses in intake i,
K_{ij} = duration of block j intake i,
L_{ij} = duration of holiday j intake i.

The total number of nurse block and holiday weeks occupied during a year = $P + Q$. As these must fit into 52 weeks the ideal number of nurse block and holiday weeks occupied per week is $(P + Q)/52$. Clearly, this

[3] Many hospitals, partly as a consequence of current financial stringencies, are examining the economic consequences of admitting three intakes a year particularly where, as often happens, the third is substantially smaller than the other two.

[4] Although in the present circumstances this approach may not be ideal in that both aspects of the allocation problem interact upon each other, it is fair to say that Banks was correct when he asserted that the design of a good *block* programme goes a long way to producing a good allocation programme. With the gradual introduction of modular schemes throughout the country the two parts combine to form only one, thus making the technique described even more relevant and adaptable.

Vol. 6
1978

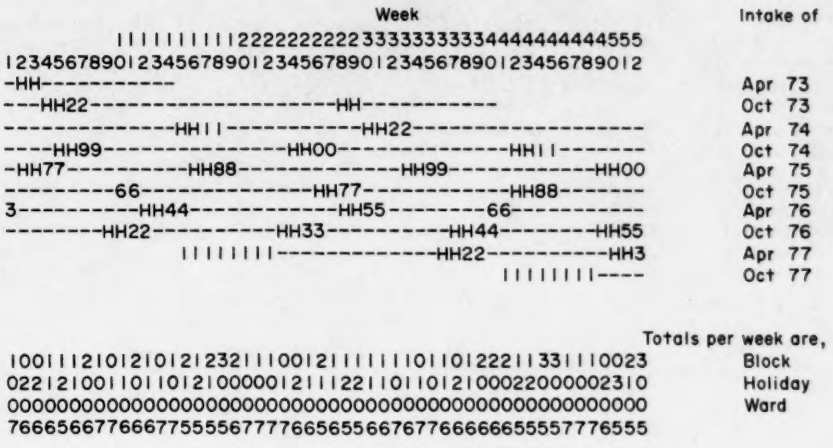


FIG. 2.

value may be fractional and the preferred number of nurse block and holiday weeks occupied per week will be the nearest integer to $(P + Q)/52$ say B and the aspiration level may be defined as $52 * B - (P + Q)$. Incidentally $(P + Q)/52$ may be rounded up or down to give B. The final solution will remain unaltered but the aspiration level will differ accordingly.

To ensure minimum allocation periods we borrow a concept from Critical Path Analysis namely that of the *dummy* activity. In order that the number of constraints is kept to a manageable size each block or holiday is considered as a block–holiday combination. This is readily achieved by adding to each separate block or holiday a corresponding holiday or block respectively of zero duration. Minimum allocation constraints can then be programmed between successive block-holiday combinations. If this strategem were not adopted then the number of constraints required would depend on all combinations of blocks with blocks, blocks with holidays, and holidays with holidays. A combinatorial problem has, therefore, been reduced to a linear one. In reality the number of dummy holidays or blocks required is usually very small due to the allocator's wish to tie many of the holiday and block periods together.

Once the constraint set has been formulated an allocation programme can be produced. As noted earlier previous documented work has been of a *heuristic* nature and has involved a single objective. As a key feature of student nurse training programmes is the interaction of the conflicting training and service considerations it was decided to solve the problem *analytically* by means of a goal programming approach. This enables the analyst to optimally solve a set of contraints where there is more than one objective to be attained. These objectives are couched in the form of constraints the right hand side of which indicate the desired level of attainment for that particular goal. The objective then becomes the minimisation of deviations from these required goals whilst still maintaining feasibility within the original constraint set. In this particular problem it is evident that the right hand sides have a particular meaning for the allocator. For the first constraint subset the goal will be not to exceed a certain number of intakes in school during each week of the year. The second subset has the 'preferred number of nurses not available for ward allocation' as its goals, and the last subset has right hand sides which are a combination of the allocator's preference and GNC requirements.

A more detailed discussion of goal programming techniques and solution procedures will be found in a subsequent paper but suffice here to report that it was found necessary to develop a new solution algorithm due to the zero–one nature of the problem and the lack of existing techniques to cope with a problem of such dimensions (a typical problem will be of the order 150 constraints and 300 variables). The new algorithm proved to be successful in coping with such large problems and is now being tested in the field to gain the reactions of allocation officers as to its practical worth. As an example of how it has been used, the weights for the problem corresponding to Fig. 1, are:

allocation period weights = $\lambda_A = 0.0097$,
ward availability weights = $\lambda_W = 0.0097$,
school capacity weights = $\lambda_S = 0.0050$.

In contrast with Fig. 2, the weights are $\lambda_A = 0.0226$, $\lambda_W = 0.0024$ and $\lambda_S = 0.0064$.

It is clear that the two programmes are quite different structurally. The set of weightings is currently developed in an *ad hoc* interactive manner with the allocator, but the ultimate intention is that they are produced in a more routine fashion.

REFERENCES

1. BANKS I (1969) Nurse allocation. M.Sc. thesis, University of Aston.
2. BARLOW AJ (1972) *Nurse allocation by computer. Nurs. Mirror* **134**(4), 43–45.
3. BARR A (1965) Scheduling of student nurses with the aid of a computer. Operational Research Unit, Oxford Regional Health Board, Report No. 7.
4. BENDALL ERD (1965) Wastage, sickness and allocation. *Nurs. Times* **61**(23), 760–763.
5. DAVIS M & SAUNDERS R (1966) Allocation student nurses by computer. *Nurs. Times* **62**(14), 467–469.
6. DAVIS M & SAUNDERS R (1966) Scheduling of student nurses with the aid of a computer. *Hospital* **62**(9), 423–427.

7. HARTMANN B (1968) Student nurse allocation. *Nurs. Times*, occasional papers 64 (1 and 2).
8. NIMMO H (1963) Student nurse allocation: Western General Hospital Edinburgh. *Nurs. Times* **59**(10), 293.
9. SAWYER J (1967) Planning for nurse training. *Nurs. Times* **63**(31), 1029–1032.
10. WHITEHEAD J (1963) Student nurse allocation. *Nurs. Times* **59**(19), 585–587.

B Moores
NW Garrod
GH Briggs
(*May* 1977)

The University of Manchester
Institute of Science and Technology
Department of Management Sciences
PO Box 88
Manchester M60 1QD
UK

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 97–98, 1978. Pergamon Press. Printed in Great Britain

Aggregate Planning with Inaccurate Inventory Data

MANAGEMENT scientists who work in both academic and industrial environments are keenly aware of the lack of acceptance of most elegant models by operating managers. This 'theory-implementation gap' has been widely studied (see references) and attributed to various factors ranging from emotional stress to poor communication. A recent attempt by the author to implement a production planning model for a medium sized plastics manufacturer indicated that management's hesitancy to adopt various models may be well founded. While gathering data required for estimation of parameters, the plant manager and the consulting team found that the information requirements of the system under consideration were not available in accurate form.

One model considered was the classical HMMS [5] LDR (linear decision rule) approach, chosen first because it is generally acknowledged to be the standard for comparison of performance for other techniques [2, p. 227]. The use of the linear decision rule for aggregate planning, however, requires accurate measurement of net inventory prior to the commencement of each production period since substantial lead time is often required to formulate plans. This measure is, in effect, a forecast which may be subject to some error resulting from a number of causes, such as inaccurate forecast of orders on hand, errors in physical inventory count, or faulty conversion of orders to inventory units. While Holt *et al.* [5, Ch. 6] have shown that unbiased forecasts of sales will yield optimal results, and inventory errors distributed with a zero mean and no serial correlation will have little effect on model performance, the production manager must consider the consequences of utilizing the LDR approach in the likely event that his own history of inventory measurement errors is not centered on a zero mean.

Knowledge of the impact of inventory errors on performance of the HMMS model may be useful in two ways. First, management can determine the sensitivity of results to inventory measurement errors, a critical consideration in the decision to utilize the model. Second, the deviation of the results from optimal performance provides an indication of the worth of a perfect inventory measurement system.

Management may easily obtain an accurate measure of net inventory by conducting a physical count and subtracting the inventory equivalent of the orders on hand on the last day of each period. A comparison of the result with earlier estimates yields a history of measurement errors. Further, the manager may correct current estimates and detect patterns which may lead to elimination of some systematic components of the error. If this is done, serial correlation in the error may be substantially reduced, and we may reasonably assume that errors in each period are independent and identically distributed.

While the analysis presented below is based on the cost function derived by Holt *et al.*[5], the sales forecast and distributions of inventory errors represent actual data modified (in accordance with the wishes of the plastics manufacturer) to represent a proportionally similar level of sales and inventory.

Sixteen separate error distributions, representative of various combinations of mean error and variance, were included in a simulation of the operation of a firm whose production and work force decisions are made by the LDR, except that inventory plus an error term generated from the distribution selected was substituted for actual ending inventory of the immediate preceding period. The sales forcasts utilized with the model were generated by the function:

$$\hat{x}_{t,\tau} = 500 + 3\tau + 115\sin(30\tau)^\circ + 55\cos(30\tau)^\circ + 40\sin(60\tau)^\circ - 80\cos(30\tau)^\circ$$

where $\hat{x}_{t,\tau}$ = a forecast made in period t for period $t + \tau$.

Each run of the simulator was continued for 120 periods, or 10 years of operation. At the conclusion of each 10-year run, both total and relevant costs for the entire period were calculated. Relevant costs were determined by deducting 'straight-time' costs from total costs and 'straight time' component of overtime from overtime costs, leaving only the 'premium' portion of overtime costs, those directly controllable by the production planner.

The results of the simulation are shown in Table 1. Note that, for very small inventory errors, the percentage deviation from optimal relevant cost is quite small. However, as the error increases, relevant costs grow quite rapidly.

TABLE 1. RESULTS

(1) μ	(2) σ^2	(3) RC_τ	(4) D	(5) $(\mu/\bar{I}^*_\tau)$
0	0	278,108	—	—
0	40.37	278,395	0.0001	0
0	186.48	278,742	0.002	0
5	40.37	278,371	0.001	0.016
−5	40.37	278,867	0.003	−0.016
10	40.37	278,858	0.003	0.031
−10	40.37	279,824	0.006	−0.031
−10	186.48	280,174	0.007	−0.031
25	40.37	283,225	0.018	0.079
−25	40.37	285,638	0.026	−0.079
50	40.37	300,337	0.074	0.157
−50	40.37	305,543	0.090	−0.157
75	40.37	329,523	0.156	0.236
−75	40.37	337,226	0.175	−0.236
100	40.37	371,240	0.257	0.315
−100	40.37	381,201	0.270	−0.315

(1) and (2) μ = mean of error distribution, σ^2 = variance of error distribution.

(3) RC_τ = total of five 10-year relevant cost figures ÷ 5.

(4) D = Relative Deviation from optimal relevant cost = $(RC_\tau - RC^*_\tau)/RC_\tau$ where RC^*_τ is RC_τ for zero error.

(5) Average inventory error divided by average optimum inventory (a measure of % inventory error) where optimum inventory results from zero error in all periods.

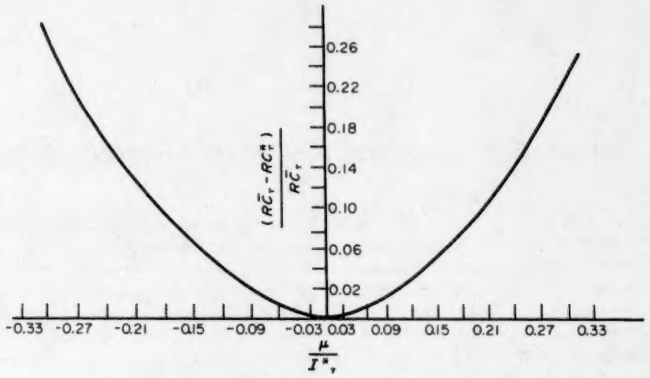


FIG. 1. *Relative deviation from optimal cost.*

The relationship between percentage inventory error and percentage deviation from optimal relevant cost is more striking and is shown in Fig. 1.

A similar examination of the effect of variance on relevant cost performance indicated a very minor fluctuation in deviation from the optimum as the coefficient of relative variation of the error distributions fluctuated widely. This finding is consistent with a conclusion by Holt *et al.* [5, p. 178] that variance of inventory errors does not drastically affect performance of the model. We may conclude that mean inventory measurement error is an appropriate indicator of suitability of the model for a particular firm. For the production planner, the implications are clear. If, for this firm, the history of inventory estimation indicates that errors are approx. 7–15% or lower, the impact on the production plan provided by the LDR may be tolerable (8% or less deviation from optimal relevant cost). However, if an inventory measurement error of approx. 30% is indicated, the impact on relevant costs is more severe (25–27%). For errors grossly in excess of 30%, it is possible that the LDR will indicate a substantial shift in the size of the work force in order to achieve rather large scale changes in planned production. If this occurs, the measure of relevant cost loses its value, and deviation from optimal total costs is likely to increase at a much greater rate.

REFERENCES

1. ARGYRIS C (1971) Management information systems: the challenge to rationality and emotionality. *Mgmt Sci.* **17**(6), B275–B292.
2. BUFFA E & TAUBERT W (1972) *Production-Inventory Systems: Planning and Control* (Rev. Ed., IRWIN RD), Homewood, Illinois.
3. CHURCHMAN C & SCHAINBLATT C (1965) The researcher and the manager: a dialectic of implementation. *Mgmt Sci.* **11**(4), B69–B87.
4. DYCKMAN TR (1967) Management implementation of scientific research: an attitudinal study. *Mgmt Sci.* **13**(10), B612–B620.
5. HOLT C, MODIGLIANI F, MUTH J & SIMON H (1960) *Planning Production Inventories and Work Force.* Prentice-Hall, Englewood Cliffs.
6. LILIEN G (1975) Model relativism: a situational approach to model building. *Interfaces* **5**(3), 11–18.
7. MALCOLM D (1965) On the need for improvement in implementation of operations research. *Mgmt Sci.* **11**(4), B48–B58.
8. RANDOR M, RUBENSTEIN A and BEAN A (1968) Integration and utilization of management science activities in organizations. *Op. Res. Q.* **19**(2), 117–141.
9. SIMON H (1965) The new science of management decision in *The Shape of Automation* Chap. 3, pp. 92–111. Harper & Row, New York.

Earle Steinberg
(*August* 1977)

Department of Management
College of Business Administration
University of Houston
Houston
Texas 77004
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 99–100, 1978. Pergamon Press. Printed in Great Britain

Switching Model for the Housing Decision

IN A RECENT article in this journal, a model is developed to assist in the housing decision [1]. Specifically, the model is concerned with the total costs of renting or buying for a given planning horizon and with the length of time necessary for the cost of renting to equal the cost of buying. The model, however, did not explicitly address the possible sequential nature of the decision. That is, for a given planning horizon, an individual may rent or buy for the entire period or he may select some 'rent–buy' combination (switch), with 'rent then buy' being the most common pattern.

The purpose of this note is to extend the existing rent–buy model to permit the consideration of the switching aspects. This is accomplished by developing a deterministic model to measure total housing costs for a given planning horizon. Costs are expressed in terminal value dollars, and it is assumed that the decision pattern is rent then buy with the buy decision being irrevocable. Similar to the existing model, the switching model is constructed to consider such factors as house values, financing terms, rental terms, maintenance costs, selling costs, taxes, and opportunity costs. With the exception of contractual financing terms, the values of the model's variables are allowed to change over time.

The formal model is presented below. In the interest of clarity, this formulation combines many of the terms used in the previous model. Specifically,

$$C = D_{bt}\prod_{i=t-1}^{n-1}\alpha_i + \sum_{j=t}^{n-1}\left(A_{bj}\prod_{i=j}^{n-1}\alpha_i\right) + A_{bn} - \Gamma_{bn}, \qquad t = 1;$$

$$D_{r0}\left(\prod_{i=0}^{n-1}\alpha_i - \prod_{i=t-1}^{n-1}\alpha_i\right) + \sum_{j=1}^{t-1}\left(P_{rj}\prod_{i=j-1}^{n-1}\alpha_i + U_{rj}\prod_{i=j}^{n-1}\alpha_i\right) + D_{bt}\prod_{i=t-1}^{n-1}\alpha_i + \sum_{j=t}^{n-1}\left(A_{bj}\prod_{i=j}^{n-1}\alpha_i\right) + A_{bn} - \Gamma_{bn}, \quad 2 \le t \le n-1;$$

$$D_{r0}\left(\prod_{i=0}^{n-1}\alpha_i - \prod_{i=t-1}^{n-1}\alpha_i\right) + \sum_{j=1}^{t-1}\left(P_{rj}\prod_{i=j-1}^{n-1}\alpha_i + U_{rj}\prod_{i=j}^{n-1}\alpha_i\right) + D_{bt}\prod_{i=t-1}^{n-1}\alpha_i + A_{bn} - \Gamma_{bn}, \qquad t = n;$$

$$D_{r0}\prod_{i=0}^{n-1}\alpha_i + \sum_{j=1}^{t-1}\left(P_{rj}\prod_{i=j-1}^{n-1}\alpha_i\right) + \sum_{j=1}^{n-1}\left(U_{rj}\prod_{i=j}^{n-1}\alpha_i\right) + U_{rn}, \qquad t = n+1,$$

where:

C = terminal value of total housing costs (dollars);
C_{bt} = downpayment and closing costs (dollars) in period t;
D_{r0} = initial rental deposit (dollars);
A_{bt} = principal repayment, maintenance costs, and after tax interest and property tax payments (dollars) at end of jth period;
P_{rj} = rent payment (dollars) at beginning of jth period;
U_{rj} = maintenance and related rental costs (dollars) at end of jth period;
Γ_{bn} = after tax capital gains less selling costs (dollars) at end of period n;
α_i = after tax opportunity cost compounding factor (decimal) at end of ith period;
n = number of periods in planning horizon;
t = period in which the switch from rent to buy is effected such that t is the first period of the buy alternative.

An inspection of the model reveals that it contains four basic components. The first and fourth represent the cost of buying and renting respectively over the entire planning horizon. Essentially these components are those comprising the existing rent–buy model. The middle two components of the switching model measure the cost of renting and buying combined.

A cost curve may be created by varying the length of time devoted to renting (i.e. varying t) while all other factors are held constant. Depending on the environment under consideration, the cost curves may take on a variety of shapes. They may be generally upward sloping, downward sloping or U-shaped. In all cases, however, there is a discontinuity (decided drop) in the curve from period n to $n + 1$.

To provide additional insight into the nature of these cost curves, a typical U-shaped curve is presented in Fig. 1. The point to the extreme left indicates the terminal value of the cost of buying for the entire planning period; conversely, the point to the extreme right depicts the terminal value of the cost of renting for the entire period. All points lying between these two extrema are the costs associated with some combination of renting and buying. Of particular interest is the global minimum of the curve; it is this point that signals the optimal switch from renting to buying.

In conclusion, the switching model that is developed extends the previous rent–buy model and provides an

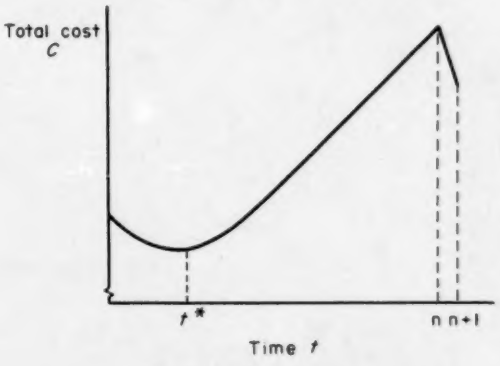


FIG. 1. *Sample cost curve.*

explicit and theoretically sound input to the common, but important, *dynamic* rent–buy housing acquisition decision.

REFERENCES

1. MOJENA R, BOOTH GG & BOWMAN JP (1976) A deterministic model for the housing decision. *Omega* **4**(1), 85–91.

GG Booth*
R Mojena
(*July* 1977)

* *Director*
University of Rhode Island
Research Center in Business and Economics
College of Business Administration
Kingston, RI 02881
USA

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 1, pp. 101–102, 1978. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

Clifton WL Jr (1970) Software Package Evaluation: An Inventory Control Example. SM Thesis, MIT, Cambridge, MA.

Coris LM (1970) Minimum Deposit Life Insurance. SM Thesis, MIT, Cambridge, MA.

De La Vallee Poussin DC (1970) Stability in Systems with Quadratic Criterion Function. SM Thesis, MIT, Cambridge, MA.

De Nervaux O (1970) Design of the User Interface for a Computerized Portfolio Selection Model. SM Thesis, MIT, Cambridge, MA.

Dreizler FD (1970) Financing Policies and Practices of Selected Urban Mass Transit Systems. SM Thesis, MIT, Cambridge, MA.

Dudnikov EE (1970) Short-Range Planning for Industrial Enterprises. SM Thesis, MIT, Cambridge, MA.

Ebeling HW Jr (1970) EDP Staffs as Change Agents. SM Thesis, MIT, Cambridge, MA.

Ellis FE (with **Ferraioli C**) (1970) The Introduction of the Computer into a Library System. SM Thesis, MIT, Cambridge, MA.

Emo Capodilista G (1970) Group Decision Theory. SM Thesis, MIT, Cambridge, MA.

Fitch JL (1970) A Study of Competition Between Intra-Corporate Operating Units. SM Thesis, MIT, Cambridge, MA.

Flautt RJ (1970) Earnings Announcements: Timeliness, Information Content and Market Price Relationships. SM Thesis, MIT, Cambridge, MA.

Fletcher PB (1970) The Design of a Program to Reduce Hourly Employee Turnover. SM Thesis, MIT, Cambridge, MA.

Frashier GE (1970) An Analysis of Factors that Affect the Share Price of Cabot Corporation Common Stock. SM Thesis, MIT, Cambridge, MA.

Garretson WW (1970) A Forecasting Model for Raw Material Control. SM Thesis, MIT, Cambridge, MA.

Goldman FW (1970) Quadratic Programming Model Implementation for Determining Freshman Scholarship Offers. SM Thesis, MIT, Cambridge, MA.

Golovin LB (1970) Two Mathematical Models for Oil and Gas Deposition. SM Thesis, MIT, Cambridge, MA.

Goodwin BZ (1970) The TFX Hearings: A Study in Controversy Creation. SM Thesis, MIT, Cambridge, MA.

Gritton DG (with **Vaughn EA Jr**) (1970) Private Sector Resources and the Urban Solid Waste Problem. SM Thesis, MIT, Cambridge, MA.

Hebeler HK (1970) Cost and Schedule Uncertainty and Prediction in Development Projects. SM Thesis, MIT, Cambridge, MA.

Heffernan JJ (1970) The Effects of Interactions in Production–Distribution Systems. SM Thesis, MIT, Cambridge, MA.

Hellman JJ (1970) Privacy and Information Systems—An Argument and an Implementation. SM Thesis, MIT, Cambridge, MA.

Hoffman CG (1970) Job Scheduling in a Multiprogrammed Environment. SM Thesis, MIT, Cambridge, MA.

Howard JJ III (with **Miles LM**) (1970) A Comparative Analysis of Attitudinal Differences Toward the Work Situation Between Unionized and Non-Unionized Public School Teachers, Data processing Professionals and First Line Industrial Supervisors. SM Thesis, MIT, Cambridge, MA.

Johnson C III (1970) Merger Activity and Investor Preferences. SM Thesis, MIT, Cambridge, MA.

Jones SD (1970) Technological Forecasting and Long-Range Planning. SM Thesis, MIT, Cambridge, MA.

Jordan JW (1970) The Two-Year College in Theory and Practice. SM Thesis, MIT, Cambridge, MA.

Jourdan BJ (1970) Feasibility of Service Time Estimates in Outpatient Scheduling Rules. SM Thesis, MIT, Cambridge, MA.

KARKANEN KM (1970) A Model Based Approach to Operational Job-Shop Control Systems. SM Thesis, MIT, Cambridge, MA.

KELLY TJ (1970) The Dynamics of R&D Project Management. SM Thesis, MIT, Cambridge, MA.

KLEINER A (with QUINT SD) (1970) An Integrative Model of the Personal Growth Process in a T-Group. SM Thesis, MIT, Cambridge, MA.

KRELL RA (1970) The Use of Interactive Decision Models by Marketing Management. SM Thesis, MIT, Cambridge, MA.

MEHTA CR (1970) Time Sharing with an Explicit Priority Queuing Discipline. SM Thesis, MIT, Cambridge, MA.

MERMEL SJ (1970) System Design for a Computerized Stock Exchange. SM Thesis, MIT, Cambridge, MA.

MITCHELL C JR (1970) NAB-Jobs in Detroit: A Black Point of View. SM Thesis, MIT, Cambridge, MA.

MOHR BF (1970) Performance Evaluation Model for Investment Analysis. SM Thesis, MIT, Cambridge, MA.

MURTHY SKR (1970) A Response Model and Heuristic for Planning Promotion. SM Thesis, MIT, Cambridge, MA.

NEVIN JP (1970) Design of a Touch–Tone Based Medical Information System. SM Thesis, MIT, Cambridge, MA.

NORDLINGER RH (1970) Some Hardware Sales Implications of IBM System 360 Software Development. SM Thesis, MIT, Cambridge, MA.

PERNA F JR (1970) An Analysis of Learning Profiles of Managers. SM Thesis, MIT, Cambridge, MA.

PERRITT HH JR (1970) Decision Making Strategies in a Political Campaign Organization. SM Thesis, MIT, Cambridge, MA.

PETERS JR (1970) A Case Study of Research and Development Communications in a Small Chemical Firm. SM Thesis, MIT, Cambridge, MA.

PETERSON DB (1970) The Use of Simulation for Short-Term Job Shop Control. SM Thesis, MIT, Cambridge, MA.

PEW RA (1970) Parity Movement Forecasting by Financial Managers. SM Thesis, MIT, Cambridge, MA.

PLOVNICK MS (1970) The Effects of Individuals' Cognitive Abilities on Occupational Choice and Behavior. SM Thesis, MIT, Cambridge, MA.

PRELLER EF (1970) Human Attitudes and Management Development in a Fast Growing Multinational Company. SM Thesis, MIT, Cambridge, MA.

PUREKA PT (1970) Evaluating Effectiveness in a Government Agency: A Case Study of the Commonwealth of Massachusetts Rate Setting Commission. SM Thesis, MIT, Cambridge, MA.

ROBINSON TA II (1970) Management Structure as a Function of Size in Process Industry. SM Thesis, MIT, Cambridge, MA.

SAUER FJ (1970) TEMPS—A Management Development Simulation for the Temporary Help Industry. SM Thesis, MIT, Cambridge, MA.

SHANE JD (1970) The Design and Implementation of a New Form of Security Trading. SM Thesis, MIT, Cambridge, MA.

SHWIMER J (1970) Single Machine Sequencing with Tardiness Penalties. SM Thesis, MIT, Cambridge, MA.

SWANSON RA (1970) A Model for New Product Attribute Selection. SM Thesis, MIT, Cambridge, MA.

TELESIO DI TORITTO P (1970) The Experience of European Direct Investment in the United States. SM Thesis, MIT, Cambridge, MA.

TODD RC (1970) A Planning Model for the Multi-Stage Production System. SM Thesis, MIT, Cambridge, MA.

VINAS R (1970) The Use of Multidimensional Scaling for New Product Positioning Decisions: An Application in the Ethical Drug Field. SM Thesis, MIT, Cambridge, MA.

WALKER RC (1970) Management Education: Corporate Activities. SM Thesis, MIT, Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 103–108
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0103/$02.00/0

EDITORIAL

Better than the Oracle?

WE ARE constantly urged to take a positive attitude to planning. But senior executives often pay only lip service to honest endeavours by their minions to grapple with difficult problems concerning the future activities of their companies, with carefully compiled plans, often brushed aside as irrelevant exercises. This is not only a manifestation of inherent contempt towards an important managerial function, but is obviously a waste of good talent inhabiting planning offices, talent that can perhaps be better employed elsewhere. This phenomenon is particularly prevalent where long range planning is undertaken, and where a typical reaction of top executives may be summarized as "We cannot possibly tell what is likely to happen in 6 months time, let alone in 5 years, hence all this planning is a bit academic".

This attitude is unfortunate and decidedly short-sighted, but understandable: a plan can only be as good as the information fed into it, and while the data on past or current operations are often accurate and reliable, the forecasts about the future may be neither, particularly where the more distant future is concerned. The dynamic nature of market trends, the cataclysmic effects of political and economic events, the rapid rate of technological change, the advent of new products and production methods—all contribute to an overwhelming sense of uncertainty, and if the forecasts on which a planning exercise is based are thought to be questionable, what credence can senior executives give to the results?

This is why so much attention has been focused in recent years on forecasting techniques, and corporate planners have been quick to point out the immense benefits likely to accrue from a relatively modest investment in this area. Clearly, anything that can be done to improve the accuracy and reliability of forecasts will greatly enhance the validity of the planning process, and that, in turn, will reduce uncertainty, will improve the ability of managers to schedule future activities, and will rebound on the well-being of the enterprise.

Forecasting is, of course, an old problem. Men have always sought the advice of fortune-tellers, and some 2500 years ago the ancient Greeks knew just the place to go to for answers to the quest of the unknown: "This place is Delphi, the centre and navel of the earth; and here Apollo prophesies to mortal men, chanting continually from his holy seat oracles concerning what is and what is to be" (Euripedes, Ion). Suppliants approaching the Oracle for a glimpse at destiny were rewarded by inspired, though often ambiguous, guesses about the outcome of future events, and from all accounts it appears that such pronouncements were taken rather seriously.

Other Delphis have come and gone, and it must have been some comfort to the followers of the management literature to discover that the spirit of

Delphi has been sufficiently revived to become a recognized concept in the parlance of modern management practice. The resurrection of Delphi in the early 1950's took the form of an exercise, named the 'Delphi project', which was carried out by the RAND Corporation of California, involving a probe into the unknown by a particular way of canvassing expert opinion. Since then the method has been widely applied, especially in the field of technological forecasting, and although it is often associated in people's minds with future innovations and R&D, there is nothing in principle that needs confine it to the field of technology. In fact, one of RAND's early exercises attempted to determine what a Soviet strategic planner would regard as the best industrial targets in the US, and to estimate the number of A-bombs that would reduce the munitions output to a given level. Similarly, the method can be used for economic forecasting development of existing and new markets, predictions of social trends, or assessments in the political arena.

What is the secret of this magic tool? Several accounts of the method and its applications have appeared in the literature; the two references at the end of this editorial are good examples, and each carries a list of further references for interested readers. Briefly, the purpose of the Delphi method "is to obtain the most reliable consensus of opinion of a group of experts. It attempts to achieve this by a series of intensive questionnaires interspersed with controlled opinion feedback" [1, p. 458]. As for technological change, "the corporate planner should pursue new means of technological forecasting to ensure his company's share in tomorrow's market" and this prompted several leading enterprises to provide a "prediction of likely inventions, specific scientific refinements, or discoveries in technology, including new applications or products which may become possible" [2, pp. 68–9].

Vol. 6
1978

Now, you can always assemble a group of experts round a table and pop the question: "When is event X most likely to occur?" or, "What is the probability that it will come about before 1990?" and so on. The group may then have what is akin to a brain-storming session and come up with some answers, but the proponents of the Delphi method fear that such group discussions may affect and distort personal opinions. As they put it, "our approach permitted us to circumvent group activity altogether and to deal with each member of our teams of predictors directly and individually. And thus we were able to eliminate the negative factors associated with group action—for example, the 'aggressive expert' who feels called upon to defend his publicly stated opinions, the senior executive with whom subordinates are reluctant to differ, and the silver-tongued salesman who can sell refrigerators to Eskimos" [2, p. 70]. "Direct confrontation.... all too often induces the hasty formulation of preconceived notions, an inclination to close one's mind to novel ideas, a tendency to defend a stand once taken, or, alternatively and sometimes alternately, a predisposition to be swayed by persuasively stated opinions of others" [1, p. 459]. These pertinent comments form the basis of the Delphi methodology, and their implications need to be fully understood when its validity and effectiveness are to be evaluated.

But, first let us describe briefly the TRW exercise [2], which undoubtedly served as a model to others over the past decade. Regarded as "probes of the technological future", the project aimed first at eliciting from panelists a list of probable technical events, in certain defined categories, that "could have a significant impact on the company". Each panelist was also instructed to assess every anticipated event with respect to three main factors:

desirability (considered from the viewpoint of the customer), with ranking categories as 'needed desperately', 'desirable', or 'undesirable but possible';

feasibility, being 'highly feasible', 'likely', or 'unlikely but possible';

timing, being the year by which the event will have occurred with three alternative probability values: 10% ('reasonable chance'), 50% ('most probable date') or 90% ('almost certain').

The next round involved a questionnaire for the list of events, collated after the first round, for each panelist to state his own assessment of the probability of every event; each respondent was further asked to indicate whether his familiarity with the subject matter was fair, good or excellent, although it is not clear whether, and if so how, such additional information can be used for weighting the estimates made.

The results of such an exercise are bound to be scattered over a wide range and the next stage is to attempt to reduce the divergence of views, either by a series of discussions with individual panelists [2] or by asking each to respond again to the questionnaire after giving him further information, such as specific relevant data requested or furnished by some of the experts in the course of the previous round. Armed with such new information, which need not necessarily include an explicit statement regarding the estimates of the other panelists, the respondent may modify his earlier assessment in the knowledge that his identity will remain concealed from the others and that no penalty will be attached to his change of mind. In many Delphi projects the feedback in each round does explicitly list the full results of the previous round, without revealing the identities of individual respondents, so that each panelist can judge for himself how he stands in comparison to other expert opinions before he is asked to react again. If he finds that he broadly agrees with the mainstrean views of the panel, he may have no reason to modify his own drastically at the next round, but if he discovers that he is very much on a limb, he may have to think again.

It is not surprising, therefore, that as the experiment continues, the estimates begin to converge. Indeed, this is the object of the exercise, to obtain a consensus and to eliminate or isolate extreme views. RAND was quite specific about this: "If the purpose of the experiment is the estimation of a numerical quantity .... it may be expected that, even if the views expressed initially are widely divergent, the individual estimates will show a tendency to converge", and added "this is almost inevitable in view of the progressively more penetrating analysis of the problem, achieved partly by means of the procedural feedback described above" [1, p. 459]. Furthermore, the RAND experimenters were keen to reduce the terminal disagreements, which still existed after five rounds, and applied their own correction procedure to modify the final results in the light of "a careful analysis of the responses", taking into account a consensus on basic assumptions, the reactions of individuals to changes in assumptions and their views on functional relationships between the major variables in the system.

The results of the RAND experiment, regarding estimates of bomb requirements and involving seven experts, are shown in Fig. 1, and clearly demonstrate the rapid convergence early on in the process, with four middle-of-the-road views remaining relatively static while the other three progressively moved toward the middle. It is interesting to note that in this case the fifth response did little to reduce the range of estimates generated by the fourth, with the ratio of the largest estimate to the smallest remaining at just over

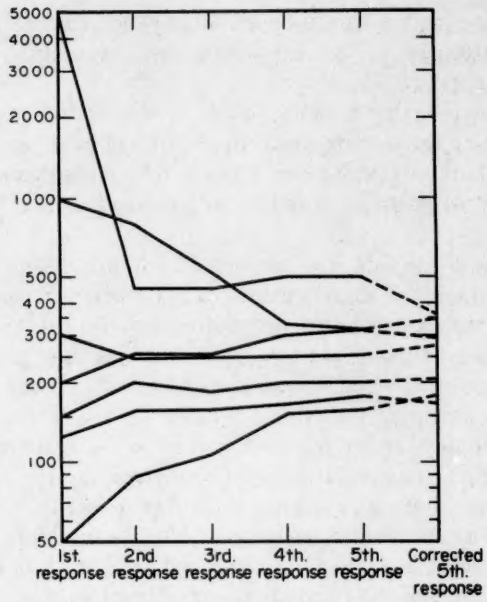


FIG. 1. *Successive estimates from a Delphi process* [1].

3:1, although the experimenters contended that there were "strong indications" to suggest that had the exercise continued, the ratio would have dropped to 2:1 or even less, and the final corrections that they applied to the fifth round of responses indeed resulted in such a ratio.

Whether the experiment did terminate prematurely, as the authors argued [1, p. 467], and whether further convergence would have taken place, is a matter of conjecture (the evidence in Fig. 1 is far from conclusive on this issue). Some measure of consensus, compared with the initial responses, undoubtedly emerged from this sequential process, though the range of terminal estimates was much wider than the experimenters hoped for or were aiming at (even after their correction procedure, which can easily raise certain fundamental objections), and it is doubtful whether the attainment of a final single estimate (if that were possible) would necessarily strengthen its validity.

Reading through the accounts of Delphi projects one cannot help but sense the implicit assumption that the truth about the future is lurking somewhere, that our groping in the dark is not an indication of inaccessibility but of ineptitude, that well-designed probes and a well-thought out search methodology are bound to get us close to and eventually even home on our target. If we do not get there, so the argument seems to go, it is simply because our methods are not refined enough; with more experience in the use of Delphi and greater sophistication in the design of feedback in our sequential probing, we should be able to develop the filter mechanism that will shut out the noise and allow us to listen to the sweet music of truth.

In fairness, it must be said that not all Delphi disciples take such a simple-minded view, or are so explicit in their expectations. The RAND experimenters were modest enough to argue that "even though the predictions obtained in the form of an opinion consensus may be lacking in reliability" they hoped that it "would often turn out to be an acceptable substitute for direct empirical evidence when the latter is unavailable" [1, p. 467]. And clearly, if members of an expert panel are heavily involved in the decision making process, or are in a position to influence it, they are more likely

to accept the results of a forecasting exercise, in which they have participated, as a plausible assumption for corporate planning, than forecasts foisted upon them by some external agents. This is pragmatism, not a blind search for the ultimate truth.

But the faith of others is far less inhibited. In their own comments on the Delphi method [2, pp. 77–82] the editorial staff of the Harvard Business Review describe it as "a method of cybernetic arbitration", whatever that means, and assert: "It seems truistic to argue that if the opinion of one expert on an uncertain point is good, the opinion of many experts will be even better". From what depths of human logic is this truism mined? They go on to say: "More often than not, however, experts disagree with one another's opinions, if not in substance, then in detail or perspective. If the opinions of a group of experts are to be put to practical use, they must somehow be boiled down into a single 'group opinion'". The single opinion, as already mentioned above, may be neither appropriate nor attainable; in any event there is no *a priori* reasons for it being *better* or more reliable than the view of the individual expert. It may be argued that it is 'better' in the sense that when a panelist changes his response as a result of feedback and knowledge of how his fellow-experts have reacted, he does so in the light of that new information, which he lacked earlier on, and that expert opinion given from knowledge is preferable to that given from a state of ignorance. This may well be true, and the results are likely to be more *acceptable* to all concerned, though they need not necessarily constitute more accurate predictions. It is quite possible for an expert under these conditions to be cowed into submission by sequential feedback (even under the protection of anonymity), abandon what may turn out in retrospect to have been an excellent forecast, and gravitate towards a common median, which is nevertheless far off the mark. Experiments carried out by Churchman demonstrate how difficult it is even for an expert who knows the right answer to communicate his views and influence his colleagues.

Indeed, there is no evidence for the magic power of Delphi to elicit the truth. One example concerning a past event is cited of expert respondents asked to estimate the number of oil wells in Texas in 1860, a figure that none of the respondents could have known, and the final median "moved close to the actual figure" [2]. How are we to understand this result? I contend that there are many explanations that need to be considered before we adopt the preposterous hypothesis that it is simply inherent in the Delphi methodology that the right answer, or something akin to it is bound to be unmasked. Instead, the example of the oil wells in Texas moved some to believe that "Delphi can and is being used to investigate matters of opinion about the historical past as well as the future, also with good results" [2, p. 82]. Assuming that 'good results' refer to the accuracy of the estimates made, should we then substitute the Delphi method for the conventional study of history? Is this the way to the promised land?

No. The major attribute of Delphi is not that it is more accurate or more reliable than other methods, but that it is an effective and inoffensive way of producing a consensus (though not necessarily a single view), and as such it is acceptable. This is nonetheless an important attribute, with generally beneficial effects on communications in a managerial system and on the adoption of a positive attitude to the planning function. But one should not expect from Delphi more than it is capable of delivering.

"The Oracle at Delphi answered questions about the future, but the replies were garbled". The followers of the new Delphi believe that the answers

of the experts "are the best answers available, small flames illuminating an otherwise cloudy, dark and unexplained future" and may therefore be regarded as the true "unambiguous oracle" [2, p. 76]. Ungarbled and unambiguous they may be, but they need not be right.

SAMUEL EILON
Chief Editor

REFERENCES

1. DALKEY N and HELMER O (1963) An experimental application of the Delphi method to the use of experts. *Mgmt Sci.* **9**, 458–467.
2. NORTH HQ and PYKE DL (1969) 'Probes' of the technological future. *Harv. Bus. Rev.* **47**(3), 68–82.

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 109–115
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0109/$02.00/0

FEEDBACK

Structural Determinism

IT IS encouraging to see Professor Eilon's editorial on Structural Determinism whittling down the part played by technology to more realistic proportions [4]. Lex Donaldson's critique [3] had already ably reviewed the whole question.

What Donaldson was not able to do was to draw attention to the crucial, indeed culminating evidence presented almost simultaneously by Blau and his colleagues [1]. Hickson *et al.* [5] and Child and Mansfield [2] had subjected the Woodward thesis [6] to a much more thorough empirical test than she and her colleagues were able to give it in their pioneering days, and Child and Mansfield confirmed the severe reservations Hickson *et al.* had placed on presumptions of technological determinism. Blau *et al.* have now reaffirmed this once again, but this time with American data.

On British firms, Hickson *et al.* argued from their results that technology was important to organization mainly at the level of shop floor arrangements, and insofar as these are reflected in the numbers of personnel in functions such as inspection and maintenance; and that technology was of relatively little significance to the patterns of organization and management activity at higher levels of hierarchy. Blau *et al.* with very similar concepts and measures clearly support this in American firms. They do so with data on no less than 110 New Jersey manufacturing concerns, many more than the 31 units covered by Hickson *et al* and the 82 covered by Child and Mansfield. Thus the results obtained from these two British studies are not simply an oddity of one particular industrial country.

This does not mean that technology has no importance. It is crucially important to what happens in and around production operations as such, as the socio-technical tradition of analysis stemming from researchers at the Tavistock Institute of Human Relations has demonstrated again and again. But there should no longer be misleading assumptions that technology plays some overriding role throughout the entire organization, and in this sense the Woodward thesis (whether it was really hers or whether it was exaggerated by others beyond her intentions) can be considered to have been refuted.

REFERENCES

1. BLAU PM, MCHUGH FALBE C, MCKINLEY W & TRACY PK (1976) Technology and organization in manufacturing. *Admin. Sci. Q.* **21**(1), 20–41.
2. CHILD J & MANSFIELD R (1972) Technology, size and organization structure. *Sociology* **6**, 369–393.
3. DONALDSON L (1976) Woodward, technology, organizational structure and performance—a critique of the universal generalization. *J. Mgmt Stat.* **13**, 255–273.
4. EILON S (1977) Structural determinism. *Omega* **5**(5), 499–504.
5. HICKSON DJ, PUGH DS & PHEYSEY DC (1969) Operations technology and organization structure: an empirical reappraisal. *Admin. Sci. Q.* **14**, 378–397.
6. WOODWARD J (1965) *Industrial Organization—Theory and Practice*. Oxford University Press.

DAVID J HICKSON

University of Bradford
Management Centre
Emm Lane
Bradford
Yorkshire BD9 4JL
UK

EILON and I are in agreement that certain generalisations propounded by Woodward should be seen as invalid, but the basis of our criticisms are somewhat different. My argument hinged upon the failure of subsequent studies to replicate the original generalisations; Eilon has been particularly concerned to point out difficulties in the original Woodward thesis.

There are just two points of clarification I would like to make about Eilon's comments.

The first is that, as Eilon's closing quotations from my article makes clear, I have argued that none of the bivariate relationships which Woodward found have been confirmed by the studies reviewed. But I would like to make clear to *Omega* readers that I have also argued that the trivariate relationships have been disconfirmed. Indeed one of the central messages of my article was that the purported replication of Woodward by Zwerman turns out, on closer scrutiny, to be a disconfirmation of Woodward's generalisations.

The second point is to question Eilon's statements about the work of Pugh and his associates, the members of what has come to be known as the Aston School of Organisational Research. Eilon writes of the discrepancies between Woodward's results and those of Pugh and his associates that: "Such divergent views clearly suggest that.... there could be differences in definitions of what is meant by organisation structure" (p. 501). The Aston measures, although covering some aspects of organisation structure very different from those examined by Woodward, also contain a sub-set

of measures comparable to those used by Woodward. These are a series of indices subsumed by the Aston School under the heading of 'configuration', and including measures such as the span of control of the first-line supervisor. A main point of my article is the comparison of the Aston School's findings on these Woodward-type variables with the findings of Woodward and Zwerman. This careful inquiry shows that the Woodward generalisations about these variables did not replicate. Hence I cannot understand why Eilon raises again a question that my article addressed and answered. Eilon goes on to make some points about problems of definitional interdependence. Whilst his methodological strictures are unexceptionable, I fail to understand the structure of this argument in this passage, and do not see how this exemplifies problems within the debate between the Woodward and the Aston researches. The Aston studies do not suffer from the kinds of limitations which Eilon implies, and a careful and critical analysis of their results is appropriate to an evaluation of the replicability of Woodward's findings. An examination of the Aston studies together with that of Zwerman, has led me to the conclusion that Woodward's generalisations have been disconfirmed.

LEX DONALDSON

Australian Graduate School of Management
University of New South Wales
P.O. Box 1
Kensington
New South Wales
Australia 2033

EILON'S editorial expresses many comments which I have heard from a variety of sources during the past few years. The weight of argument and evidence have led to my developing similar views to those expressed by Eilon. With so many organizational variables studied it seems intrinsically unlikely that anyone (even technology) could be the sole or principal determinant of structural features. The recognition that organizations are comprised of complex, interacting and interdependent variables leads one away from any specific prescription for success; notions based upon the principle of equifinality seem currently to be more relevant than normative general solutions. In any case, the notion that structure can be the determinant of success appears impractical in ignoring both market factors and the behaviour of people, both of which can obviously affect the degree of success achieved independently of structural factors.

What really is so remarkable is that a study carried out prior to 1958 should contain conclusions which are important as a subject for editorial debate in late 1977. With the benefit of almost 20 years of intensive research into organizations, with correspondingly wide research findings and analyses by management schools, we are now in a position to disconfirm the theory postulated.

Whilst I agree with the views expressed in the editorial, I am disappointed at the omissions rather than the comments included. This study has been relevant for many years and has served a significant heuristic function for many researchers, with regard to topic area and organizational research methodology, so that it has been and is a very important study. Consequently, although agreeing with Eilon's factual comment, I consider that his obituary could have been more generous.

JOHN B CHAPMAN

The University of Aston Management Centre
158 Corporation Street
Birmingham B4 6TE
UK

I FEAR that Eilon's editorial does an injustice both to the seminal work of Joan Woodward and to the larger body of research by what some call the 'structuralists'. Woodward's research is part of that social science tradition which argues that structure must be built up from and relate to the division of labor at the operating level. As Eliot Chapple and I have sought to demonstrate, work flow serves as the primary determinant of how people tasks and departments are grouped.

The market place, governmental pressures and unions and a whole host of external factors obviously impact the structure in shaping roles which, in turn, must also be incorporated in the structure based on work flow criteria. Above the level of operating work, the major axes or structural affiliations of the organization reflect a variety of other factors, most particularly whether product, geography or functional considerations are to be paramount.

Needless to say, few would dare argue that managerial capability is insignificant in determining organizational success. No matter what the structure, incompetence is incompetence, and it is deadly. But managers can be aided in shaping structure by knowing how to relate division of labor decisions to the design of organizations, moving from technology upward.

Those who seek to watch and learn from the growing literature on organizational behavior recognize the degree to which Ms Woodward's work was the stimulus for an enormous outpouring of research on management structure. In large measure that critical variable had been ignored, left to the normative pronouncements of scientific management rather than to empirical testing. She pioneered in extending field work into these areas and stimulating a whole generation of researchers by her interests and methodology.

LEONARD SAYLES

Graduate School of Business
Columbia University
New York
NY 10027
USA

THEORY derives from a study of research based data recording observations of physical and behavioural phenomena. In relation to management both must have relevance to, and indicate developments in practice. Often the practitioner is not able to spend the necessary time in determining the relevance of theories to his own situation. The role of the management educator is to help make the transition between theory and practice through a synthesis of research findings within an appropriate framework.

Apart from the theoretical shortcomings of Woodward's work already referred to in Eilon's editorial [2] it is difficult to help managers to appreciate its practical application. This short comment is addressed to the

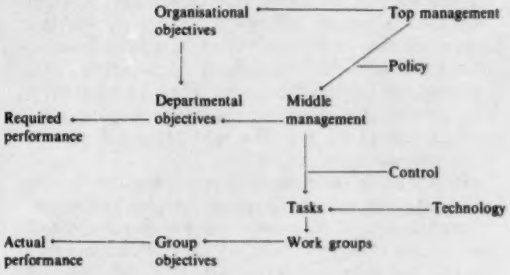
nature of these difficulties and indicates an approach which allows a synthesis within a managerial context.

If the results of Woodward's research can be accepted in a particular technology, a given span of control for supervisors seems to be associated with success. Unfortunately, the results do not indicate those units of the production process for which a particular supervisor carries responsibility. This factor can be very important in determining the success of an enterprise. Managing the discontinuities of a production process can be crucial. This can be exemplified by the relationships and control procedures which operate between production and despatch departments. A failure in the communication system between the two units results in lack of coordination with subsequent failure to despatch goods on time. The solution may be to allow one department to become dominant in managing this interface. Thus it is possible to have the same span of control with different reporting relationships but with equally successful performance. Knowledge of the span of control does not help in the design of an organisation. It is managing the interface which is important. Managers are quick to point this out.

In the so called structural approach the practicing manager gains no advantage from a knowledge of the span of control of the managing director or the number of organisational levels occurring in successful companies. What the manager seeks to understand is how the behaviour involved in the structure is reliably manifested by the participants of various work categories in specifiable circumstances. The rules governing the structure can then be altered to improve performance as the circumstances change. This behavioural pattern results from an interaction of formal and the informal structures that have developed alongside one another. Structuralists tell only part of the story.

When Woodward wrote in 1970 she acknowledged that the task and the way it is planned and controlled is important in determining organisational behaviour. It is surprising she was unaware of the work of Chapple and Sayles [1]. Their research indicated the different types of work groups associated with different production processes. One such group, named the erratic group, was characterised by an apparent lack of relationship between the seriousness of the grievance and the intensity of their protest. Such groups were found to be located at points in an assembly or material flow where there was inadequate opportunity for workers to communicate. The lack of coordination can result in loss of incentive earnings or additional physical labour for the employees. A trivial complaint is quickly changed into a major issue. This is the type of behaviour resulting from technology that managers understand. In this instance the knowledge would help them diagnose problems associated with work behaviour and take action either in changing the style of management or altering the production flow. Woodward's theory is too abstract in many of its aspects for managers to relate it to their experience.

The work of Hickson *et al.* [3] was basically on various dimensions of structure. They were aware that technology was related to those components of organisational structure located close to the actual work flow. Their research findings agree with the view of Chapple and Sayles [1]. The acid test for any theories of management is how far these can be related by the practitioner to his own activities. In addition they should help him view relationships in a different and more insightful manner. The problem with the structuralist approach is that it is skeletal rather than substantive. It is static rather than dynamic. As already indicated span of control is important. However it is more important to show how the management process works with various categories of this dimension. An example of a more dynamic model is given in the accompanying diagram.



Top management determine policy and organisational objectives. These are passed down to middle management where they are translated into departmental objectives and control systems. The tasks associated with the technology of the company affect the type of work groups which are formed. From the Hawthorne studies [4] onwards it has been known that work groups will set the norms of performance for their members and develop their own group objectives. The discrepancy between departmental objectives and the actual performance is partly a function of the control exercised by management. The greater the effective control the smaller the difference between the required and attained performance levels. By using models such as this a manager is able to trace where the problem lies, i.e. in the work groups, in the organisation of the tasks related to the technology or the control systems being operated. It may be that policy and organisational objectives are wrong. Whatever the cause such models direct the manager's attention to key areas for examination. They synthesise work from a variety of fields and can be quickly adapted to a manager's particular situation.

It is doubtful whether Woodward's theory was ever used to help in the design of an organisation. As has been indicated it is not the type of theory which would lend itself to translation into the practice of management. It therefore has few implications for the practitioner. The continued popularity of the theory has been sustained by its uncritical incorporation into the academic literature as a source of reference.

REFERENCES

1. Chapple ED & Sayles LR (1961) *The Measure of Management*. Macmillan, New York.
2. Eilon S (1977) Structural determinism. *Omega* **5**(5), 499–504.
3. Hickson DJ, Pugh DS & Pheysey DC (1969) Operations technology and organisational structure: an empirical reappraisal. *Admin. Sci. Q.* **14**(3), 378–397.
4. Roethlisberger FJ & Dickson WJ (1939) *Management and the Worker*. Harvard University Press.

Charles Baker
Durham University Business School
Mill Hill Lane
Durham DH1 3LB
UK

IN A recent issue of the Journal of Management Studies, Lex Donaldson [1] reviewed the debate surrounding Joan Woodward's work on 'technology', 'organisation structure' and 'performance'. Subsequently, in an editorial in *Omega*, Professor Eilon [2] doubts whether, following Lex Donaldson's critique, Woodward's original theory can possibly survive. It is my contention that there are many theoretical considerations that have received little or no attention in the literature on Organisations, and which are capable of throwing a different light on the relationships at issue, and the arguments used. This note draws attention to these considerations.

Several issues are at stake. Some of them are simple and easily disposed of. Others are complex and worthy of being untangled. One major issue is whether Woodward's theses (for there were several of them) were defensible and illuminating. In particular, the nature of the influence of technology on 'organisational structure' provides a focus for Woodward's critics. On this, she made no claim to *originality*, claiming only to shed light on Veblen's suggestion. We need to know, did Joan Woodward in fact take the position that she is accused of taking? A second group of problems is concerned with whether, and if so, in what way, 'technology' can be sensibly said to determine 'organisational structure' (whatever that is taken to mean), or more generally, whether what has been called 'structural determinism' is defensible. A third matter of importance is, whether or not she can reasonably be held to have taken them, do the views that her critics attribute to Woodward merit serious attention? A fourth group of issues surrounds the question: can any useful generalisations be made about the relationships between technology, 'organisation structure' and 'success'? Although justice cannot be done to all of these issues in a brief note, there are some basic observations that should be made.

It can be said with confidence that at least some of the major issues in so far as they are empirical, can be decided without the aid of a major research project, and that so far as they are not empirical, then no amount of research can settle them, or indeed have much of value to contribute. Other issues, of course require painstaking empirical research, and in fact, have received it. I shall make reference to this work only in passing, since the purpose of this note is one of conceptual clarification.

Turning to the grounds on which importance has been assigned to the issues: Woodward's critics agree that her work was centred upon the relationship between 'technology' and 'organisation structure'. Lex Donaldson's paper is very cautious, and claims to "provide a detailed analysis only of the particular thesis of Woodward" [1, p. 257], although he argues that an examination of the thesis "is neither trivial nor without implications for the wider paradigm", which I take to mean that examination of a particular Woodward thesis might well throw some light on the concepts of *technology* and *performance* to which he refers in the same passage. The particular thesis in question, as Lex Donaldson expresses it, is:

> "The core of Woodward's analysis is a scale of production technology which distinguishes nine ordered categories, which can be conflated to three main categories of: unit and small batch production, and process production .... Woodward terms this a scale of technical advance, the higher categories are seen as new and prototypical of the future and they appear to be more capital intensive or at least to be less labour intensive".

It will be observed that there are several hypotheses here, and it is some of these that Lex Donaldson claims to be 'disconfirmed', namely "the trivariate relationships between structure, technology and performance" [1, p. 273].

But are they important? An earlier observation in the same article was, "There seem to be at least two features which make Woodward's thesis so noteworthy. Firstly, the major determinant of organisational structure is imputed to the form of production technology, i.e. to a very tangible and graphically communicable social fact. And secondly, the relationship of these two variables, the fit between technology and structure is shown to be related to organisational success. This is a combination of features which is almost irresistible for both the academic and the practitioner".

In on *Omega* editorial, Eilon makes much the same point [2]. Woodward's theory "would not stand up to a searching examination. It was too simple, too static, too pat...." and Lex Donaldson's critique "has so damaged its very foundations, that it is doubtful whether the theory, in anything resembling its original form can possibly survive". He adds "it must be said that there was something naively attractive about the proposition that the organisational structure and processes of a manufacturing firm are causally related to its technology". Eilon classifies this view as 'Structural determinism', arguing that "it ignores the personalities of the top executives in the managerial system, and the possible effects that they might have on the development and exercise of control procedures: it argues instead that the organisation structure of a given firm is the result of a process of adaption in response to processes exerted by the environment."

Having established what Woodward's critics take her position to have been, it is important to note that Lex Donaldson, in his paper, is careful to say that the 'original Woodward thesis' is disconfirmed, and, not so much by the work of Zwerman, as by a detailed critique of it. To say that it is disconfirmed is not of course the same thing as to say that it is refuted. Lex Donaldson amends this position, weakening it somewhat, in his conclusions: "The plainest way of interpreting the current evidence is that it disconfirms core aspects of the Woodward thesis". [1, p. 273].

The present charges laid against Woodward, can thus be summarised: Woodward is held to have taken the position that the organisation structure, by which is meant the shape of the hierarchy, levels, ratios of supervision and so on is determined by technology, that is, that technology is the *major* determinant (technological determinism) and that business success is assured if firms obey the dictates of the determinants, and denied it if they do not, and that as a matter of fact, firms do obey the dictates. It is these theses, and the weaker, statistical thesis that there is a tendency for this to happen, and that the tendency is powerful enough to justify the entertaining of the view that technology determines structure, that are under attack. The propositions are, as has been pointed out, inconsistent. It should be noted that Woodward's remarks on these topics are frequently ambiguous and so heavily qualified that it is pointless to attempt to specify precisely which thesis she adhered to, and precision here is of some importance. What is certain, however, is that she specifically rejected the hypothesis that technology is the only, or the most important determinant of 'organisation structure'. She says [4, p. 50]:

> "Therefore the main conclusion reached through this research project was that the existence of the link between technology and social structure first

Vol. 6
1978

> postulated by Thorstein Veblen ..... can be demonstrated empirically: it is not suggested that the research proved technology to be the only important variable in determining organisational structure, or that such factors as the history and background of a firm and the personalities of the people who built it up and subsequently managed it were unimportant".

After citing some examples, she continues, [4, p. 51] "technology though not the only variable affecting organisation, was one that could be isolated for study without too much difficulty. The patterns which emerged in the analysis of the data indicated that there are prescribed and functional relationships between structure and technical demands" and "Trends in organisation appeared to be associated with an increasing ability to predict results and to control the physical limitations of production".

Thus, what she did say about the relationships between technology and 'organisational structure' was ambiguous but she did expressly preclude the interpretation which would make the expression 'prescribed and functional relationship' mean the same as 'uniquely determines', or 'the only important determinant'. In short, she expressly disclaimed the position which she is accused of taking, and what she did say was too ambiguous to be able to say with confidence which of several other interpretations, or hypotheses she did adhere to. Thus the proposition that technology uniquely determines 'organisation structure' is not one to which Woodward subscribed, although it is probably the one for which she is most widely known, and is a legitimate target for criticism. In this context it should be emphasised that Woodward's critics are not to be held responsible for the ambiguities in her texts. The proposition in question could, of course, be made analytic so that the organisation's structure is identified with its technology, but Woodward did not do that either.

I shall, in what follows, refer only to the doctrine that Woodward's critics attribute to her, since that is the one under discussion. To distinguish it from the many theses that Woodward did express, it will be called 'The Woodward Straw Thesis'. If someone did hold such a thesis it would be reasonable to ask: is it possible to reconcile the two propositions that (1) technology is the sole determinant of 'organisational structure' and (2) that firms whose technology is of the same form will be 'successful' (whatever that means), only if they adopt a particular and identical 'organisational structure' (as envisaged by both Woodward and her critics). It is possible to do so. One way would be to propose that, *in equilibrium* technology will generate a particular '*organisational structure*', or *in the long run, ceteris paribus*, it will do so, or it will do so only given a *certain degree of similarity*. This last move is one which Woodward does appear to have taken, but in respect of the weaker proposition that technology is a *major, partial determinant* of actual and ideal 'organisation structures'. It is in fact a move of the same logical type as that often made by marginalist economists in defence of the classical theory of the price mechanism, against attacks from those who argue that its assumptions are unrealistic. There are faults with this line of argument, but it is a common view, and there are no knock down arguments which can achieve a definitive refutation. More importantly, it is to be accepted or rejected on theoretical grounds.

Most of the moves available to economists are open to a holder of the Woodward Straw Thesis. That people do adopt or could adopt this line of approach is a reliable indication that the doctrine is not an empirical hypothesis at all, but is what Kuhn (and Wittgenstein) have called a *paradigm*, but which, in an older terminology would be regarded as a presupposition or a metaphysical proposition. The economic theory of markets is a doctrine which could reasonably be classified as 'structural determinist', and is regarded by many as none the worse for that. Determinism itself is, of course, a metaphysical issue, and because of that, no amount of evidence, from research projects or elsewhere is capable of dislodging the theory. Only a better theory can do that, together with the weight of evidence which is already available: the role of empirical research is then to gather more specific evidence, for more specific hypotheses.

I argue that the weight of argument and evidence are sufficient to justify the rejection of both the Woodward Straw Thesis and the conception of organisational structure to which it relates. There is a much more powerful theory which regards technology, along with other forces, such as those of the product market and the labour market as important influences on (partial determinants, together with others) the behavior of organisations and of their members. This way of thinking, which has been called 'structural particularism' [3], takes many forms, and explicitly builds in the values and policies of organisation's members, including their managers, where appropriate. Although the perspective or theory does take many forms, it could be described briefly as seeing *technology*, the *product market* and the *labour market* reacting with the *administrative arrangements* (and other infrastructures), as well as with the culture of the organisation and its members, law, the national economy, and much else. All these can be shown to generate pressures upon the members of a firm to behave in particular ways, or restrain them from doing so, as well as to provide opportunities for change, or innovation.

An important feature of this approach is that 'organisational structure' becomes a much wider and dynamic concept, identical with the particular arrangements of technology, product market, and labour market states, and of *administrative arrangements*. This conception of administrative arrangements is equivalent to the notion of 'organisational structure' used in the work under scrutiny and occupies a much lower place in the logical ordering of organisation-concepts. The matter of whether it is determined by technology is correspondingly less important. There is an abundance of detailed studies, and of practical experience, which should render the proposition that managers can change the administrative arrangements or devise new control procedures beyond dispute. A determined holder of the Woodward Straw Thesis (if there are any) could no doubt produce auxiliary hypotheses to show that these changes must have resulted from technical imperatives, but this would produce little gain. Organisations' administrative arrangements in fact derive from a variety of sources, ranging from pure whim on the part of some managers, to fashion, and to carefully designed infrastructures, although in my experience, the last is not all that common.

Probably fashion is the most frequent, but this *is* an empirical matter. Whatever the source of particular sets of administrative arrangements, it is difficult to deny, sensibly, that the results are often different from those expected and hoped for by managers, and it makes a good deal of sense to say that the actual results are partly determined by technological and other technical constraints, and pressures among others. I suggest

also that technological and product market pressures and constraints can take many forms and combinations sometimes impelling managers into some kinds of action, sometimes so constraining that no one can make any significant changes. It all depends.

These states can be mapped out, and diagnostic procedures can be devised. Structural forces do, in part determine behaviour, but on this view, both technology and the narrower conception of 'organisational structure' are both structural, partial determinants of behaviour, and part of the organisation's structure. There is no reason why interaction between these should not be postulated, and this is a useful insight discussed by the 'systems theorists'.

Perhaps it is time for the Woodward Straw thesis to be given a decent burial, and for attention to be turned to the developing art of Organisational Design, to which the concept of structure seems to be a key, and which the work of Joan Woodward, and indeed some of her critics, helped to bring into being.

REFERENCES

1. DONALDSON L (1976) Woodward, technology, organizational structure and performance—a critique of the universal generalization. *J. Mgmt St.* **13**, 255–273.
2. EILON S (1977) Structural determinism. *Omega* **5**(5), 499–504.
3. LUPTON T (1971) *Management and the Social Sciences.* Penguin, London.
4. WOODWARD J (1965) *Industrial Organization—Theory and Practice.* Oxford University Press.

JOHN DONALDSON

Department of Management Science
Imperial College
Exhibition Road
London SW7 2BX

Organisational Choice and the Woodward Thesis

"IT HAD to come" writes Professor Eilon in his editorial [2] on the latest attempt [1] to discredit the Joan Woodward thesis that the organisational structures of manufacturing firms are shaped by their technologies [6]. Why did it have to come? Why is it important that this thesis should be discounted?

Eilon reviews various methodological reasons, but his key reason for feeling it is so important to reject the theory lies in a philosophical objection to the whole notion that organisation structure may be causally determined by their technology: "Such a conclusion simply offends the intuitive belief of many of us in the existence of choice in organisation design and the validity of the notion that firms can be restructured as a result of conscious managerial decisions".

One might go further than this. Any notion that organisation structure, or any other human institution for that matter, may be causally determined by an impersonal force such as technology must surely be rejected. Not to do so is to deny the possibility of human beings shaping their own destiny, and commits one to adopting an ideological stance that is essentially conservative. To believe that organisation structures are inevitably determined by the constraints imposed by technology or size is to deny, for example, the possibility or reshaping organisations to redistribute power and make them more democratic, or even simply to reduce the size of the management overhead. It is to accept the inevitability of things as they are.

But what is the evidence for refuting the Woodward thesis in Donaldson's article [2] which prompted Professor Eilon's editorial? The article reviews various attempts to replicate Joan Woodward's original South East Essex study, and concludes that the evidence not only fails to support the thesis that technology is the prime causal determinant of structure, but in fact suggests that other determinants, in particular size, may be more important. The article does not attempt to reject deterministic arguments as such, and its criticism of the Woodward thesis is conducted in the same deterministic frame of reference. Thus, the evidence for rejecting the Woodward thesis does nothing to establish the alternative thesis of organisational choice that Eilon would like to see in its place.

Joan Woodward did not in fact advocate a naively deterministic view of organisation structure. What she found out in her South East Essex study was very simple: In the case of mass and process production at one end of the technological spectrum, and in unit production firms at the other end, certain features or organisational structure appeared to be associated both with the technology and with business success. Joan Woodward interpreted these findings as suggesting that there was an *appropriate* structure for particular production technologies, and if a firm wished to be successful it would be well advised to match its structure to its technology. In effect what she did was to replace the old concept of there being 'one right way to manage' with the idea that there was a right way to manage for different tasks. Her subsequent work was in many respects a search to identify more of these 'right ways', in particular to find right ways of structuring organisations in the batch production area for which her South East Essex research had failed to produce guidance.

However, even this weak version of technological determinism needs to be rejected on the pragmatic grounds that it still implies that choice of structure is restricted. At a time when we are trying to invent new democratic structures for industry, it would be intolerable to allow our vision to be blinkered by views of 'appropriate' structures based on research into past practices. Criticism of the deterministic thesis needs to be fundamental.

The simplest starting point may be to assume, despite the repeated failures of replication, that Joan Woodward's original South East Essex data was valid and that she did in fact find links between certain sorts of production technology, organisation structure and business success. What non-deterministic interpretation may one place on this?

It seems plausible to suggest that what Woodward found were simply the organisational solutions in vogue in the mid-1950's for administering certain production technologies, and which most progressive and successful firms had adopted. Since then other problems, for example size or overhead costs or industrial relations, may have become more pressing with the result that managers or their consultants are shaping organisation structures to meet different priorities and

Vol. 6
1978

purposes. It may be simply that fashions in organisation design have changed. Either way this would account for failure to replicate what was essentially a time-bound, and perhaps also a place-bound, study. Thus although Woodward's data showed a link to exist between technology and structure among successful firms, there was nothing in the data that demonstrated this to be a necessary link.

Had such a line of thought been pursued at the time it might have led Woodward to explore how far her data were compatible with the concept of managerial choice. For example, she might have pursued in her analysis the possibility that firms whose structure deviated from the median pattern might have had some significant lesson to offer. Some of these firms, had they been included among her follow-up case studies, might have revealed equally satisfactory alternative structural solutions for handling the same production technology —perhaps even some innovative approaches to organisation design.

It would be extremely misleading, however, if the impression were conveyed that all Joan Woodward's work was cast in a deterministic mould. During the later research by her research team at Imperial College, the theoretical emphasis of the research shifted considerably. The concept of production technology was broadened to the more embracing concept of 'task' [3] and attempts were made to identify the key aspects of production tasks with which the organisation's structure and its planning and control system had to cope—the so-called 'task analysis' approach. The essence of this approach, in common with the American 'contingency' theorists [5], is that organisation design is seen as potentially providing a 'solution' to the demands of the task. Attention is thus focused on analysing those features of the task, for example uncertainty and complexity [4] that managers would be well advised to take account of in designing their organisation[1]. This is a far cry from the determinism implicit in Woodward's earlier work. Curiously Donaldson [1, p. 258] chooses to ignore this development on the grounds that it is "separate from the 'achievements' of the Woodward tradition", though it is hard to see what the Woodward tradition might consist of other than the work she did or sponsored both earlier and later.

Attempts to undermine Woodward's work by advocating the case for alternative determinants of structure than technology are essentially sterile. The ghost of determinism in this field needs to be laid to rest once and for all, and research to focus on the nature of the choices facing management and how these choices may be resolved.

To a large extent this development has already taken place, though there is still a tendency to assume that there could be 'right' structures for particular tasks. Perhaps some constraints seem so powerful as to preclude choice, but that merely says something about our inventiveness.

[1] It is worth noting that this study [4] showed that the planning and control system adopted by many batch production firms copes with the technology in such a satisfactory way that the technology itself appeared to place no constraints on the structure, thus allowing virtually any structural arrangements to flourish. This finding could well explain Woodward's original failure to find a link between technology and structure in the batch production area.

REFERENCES

1. Donaldson L (1976) Woodward, technology, organisational structure and performance—a critique of the universal generalisation. *J. Mgmt Stat.* **13,** 255–273.
2. Eilon S (1977) Structural determinism. *Omega* **5**(5), 499–503.
3. Kynaston Reeves T, Turner BA & Woodward J (1970) Technology and organisational behaviour, in *Industrial Organisational Behaviour and Control*, (Ed. Woodward J) pp. 3–18. Oxford University Press, Oxford.
4. Kynaston Reeves T & Turner BA (1972) A theory of organisation and behaviour in batch production factories. *Admin. Sci. Q.* **17**(1).
5. Lawrence PR & Lorsch JW (1967) *Organisational and Environment: Managing Differentiation and Integration.* Harvard University, Division of Research, Boston.
6. Woodward J (1965) *Industrial Organisation—Theory and Practice.* Oxford University Press.

Tom Kynaston Reeves
(Formerly a member of the
Joan Woodward research team at
Imperial College, London)

Anglian Regional Management Centre
156/164 *High Road*
Chadwell Heath
Romford
Essex RM6 6LX

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 117–121
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0117/$02.00/0

On Optimal Control and Macroeconomic Policy[1]

MAURICE H PESTON

University of London

(*Received September* 1977)

Control theory has important contributions to make to the theory of macro-economic policy. These include a unified view of estimation, forecasting and control and a consequently clearer focus in identifying performance criteria. Further, the methods of control theory imply objectivity and a need for assumptions to be made explicit. By contrast, the UK lacks a research body for forecasting and policy evaluation detached from doctrinal or methodological commitment. Engineering control methods, however, diverge from those wholly acceptable to economic systems in that, in the latter, the controller is himself part of the system he controls and some of his controls operate only indirectly. A further difference is that economics cannot as yet furnish the controller with reliable models of system behaviour linking policy instruments to set targets.

ALTHOUGH there are certain clear cut and important differences between the formulations of control theory and those of the theory of macroeconomic policy, this does not, in my view, detract from the usefulness of control theory to the economist. Let me start, therefore, by listing the benefits the economist can gain from an appreciation of control theory, but then go on to point out why the two formulations are not logically identical. Essentially what I shall be saying is that it is useful for the economist and others interested in economic policy to know something about control theory, but that is not all they need to know!

THE CONTRIBUTIONS OF CONTROL THEORY TO ECONOMICS

In its most complete form the control theory approach involves a unification or simultaneous determination of estimation, forecasting, and control. It makes clear that all three are subject to the same principles of optimality, and are part and parcel of the same problem. Of course, in particular cases they may be separable. The best known example of this is where the optimum forecast may be obtained on its own, so to speak, and compared with the desired outcome, the optimum control being derived subsequently. But it is as well to bear in mind that such separability is not always possible. The obvious examples concern problems in which the utility function is of a higher degree or of a totally different form from the quadratic and in which the constraints are non-linear.

There is the simple point that an optimal estimate or an optimal forecast depends on the purpose one has in mind. In the policy context this implies that it is the advantages that accrue to those that make the policy, or to those they are making it for, which comprise the criteria to be applied. Curiously enough, all this should be obvious on the basis of elementary microeconomic theory. Nonetheless, in practice in econometrics and in economic forecasting it is sometimes neglected. Another way of putting the point is that the criterion of what is a good estimate or a good forecast is too much taken for granted, and not related

[1] This paper was basically written before I took up my present post as Senior Special Adviser to the Secretary of State for Prices and Consumer Protection in the UK. Nevertheless, it is worth emphasizing that nothing contained therein is to be interpreted as official policy.

explicitly to the objectives being pursued. It is also to be noted that those who are specialists in the theory and practice of economic policy do not pay as much attention as they might to criteria of performance. They hide too easily behind the difficulties of discovering what the preferences of decision makers are. The result is that despite the advances in recent years in estimating behavioural or structural relationships, we have neither equivalently satisfactory estimates of government reaction functions nor a format for ascertaining such functions in the future.

THE ROLE OF SIMULATION

This leads to the consideration of simulation as a necessary concomitant of the control approach. I see this as made up of two parts:—

(i) the investigation of an estimated model to see what its behavioural possibilities are;
(ii) the control of the estimated model to check the sense and more broadly the acceptability of the performance criteria.

The first of these is not unique to the control approach. It is merely that control theorists have taken it rather seriously, and economists ought to take a lead from them. In particular, it is vital to make the mental leap involved in accepting that the economic history of any period is not the only possible history of that period. (I might add parenthetically that there is a serious philosophical problem implicit in what I have just said. There are theorists who regard what happened as what had to happen. They are obliged to reject as meaningless simulation in the sense that I understand it. There are other theorists who emphasise the stochastic nature of the problems we are discussing, and, therefore, the existence of limits to what can be explained in practice. Nonetheless, even they have great difficulty in refraining from putting forward specific causes of everything that occurs. In sum, there are few economists who in practice are not determinists—even naive determinists).

Too frequently economists estimate models, especially big models, without then devoting a great deal of effort to exploring their behaviour. The purpose of model building appeared at one time to be little more than providing a basis for yet bigger model building. The result is that we have no shortage of economic dinosaurs to place in a museum of the history of economics if we were ever to have one.

Happily that state of affairs is now changing rapidly. In this context my own experience of simulation methods at PREM was quite fascinating. Even the small, and extremely simple, model that we were manipulating showed remarkable behavioural variations. What to the naked eye looked to be fundamentally different time paths were generated from the same initial conditions, the same economic structure, the same policies, and the same stochastic processes. Our danger was not that of believing that only one history of the economy was possible, but rather that any history appeared to be quite arbitrary.

Turning to diagnostic checks on the objective function, it is important to emphasise right from the outset that in a world as complex as ours is, it is wrong to assume that there is a true underlying set of criteria which can be reached by some form of psychological analysis of politicians or senior officials. They may know what they like when they see it, but it is absurd to assume that there is any direct and easy route to ascertaining their preferences or expressing them as utility functions. Thus, what we are really trying to do is two things:

(1) to discover how different ways of controlling a particular model using a particular criterion function lead to variations in economic performance. In particular, one seeks to discover whether some variations in the criteria are critical and others are not in determining the output of the system;
(2) to discover how using particular criteria to control particular models leads to actual policies. In particular, one is anxious to discover what circumstances are critical for what types of policy.

CLOSED AND OPEN LOOP SOLUTIONS

The next topic I would like to raise concerns the distinction between open loop and closed loop solutions to policy problems. An extraordinary amount of economic optimisation

theory is in a form which yields open loop solutions. Systems are optimised from now to infinity yielding an infinitely long path for the policy instruments. Even in explicitly stochastic models, because of the certainty equivalence theorem, a similar conclusion pertains.

My own view is that the greater emphasis should be placed on the determination of closed loop solutions relating the new input of data to variations in policy in the next few periods. Apart from anything else such an approach is *prima facie* much more economical than any other, and should therefore appeal to members of the economics profession on that account alone. It is also something that ministers can understand.

The suggestion is sometimes made that the closed loop approach makes economic policy automatic, and policy makers mere robots. This is erroneous. If the world were entirely fixed in structure and we knew all there was to be known about it (including its stochastic aspects) such automaticity would emerge. But in practice we are always investigating and learning; thus, plenty of room remains for the use of judgement. What is most important is that the judgement should be explicit and capable of being distinguished from the model itself. This, indeed, leads me to my main recommendation.

THE NEED FOR OBJECTIVITY

The justification of the optimal control approach to economic policy (and of some other approaches, for that matter) is that they make explicit all their assumptions, including those of method, and the data they employ. They are objective in the sense that experiments in the natural sciences are objective. This does not mean that no judgement or feel for the problem is involved. Quite the contrary, a great deal of skill, expertise, and experience is required. But it does mean that in principle the work is repeatable and transferable. There exists a recipe, so to speak, even though a good cook will succeed with it where a bad one fails.

What is lacking in the UK at the moment is any such objective approach to economic policy. None of the forecasting models in use have the desired characteristics, and economic policy itself is not decided in that way. The result is that the assessment of policy is always blurred, it being difficult, if not impossible, to disentangle the subjective, personal element from the rest.

Now, I hasten to add that I am not arguing that the personal element should be left out of policy making. And it is apparent that there is always information accruing to the decision maker which is not easily encompassed in a model. I am arguing that if models are to be improved, and if the significance of judgement itself is to be appreciated, the two must be distinguished.

Anyone with experience of models is aware of the seeming nonsense they generate on occasions. This or that forecast makes no sense. In due course we discover this or that peculiar policy is suggested as optimal. Nonetheless, despite the fears of those who are unduly sensitive to public criticism from the ignorant, such objective work is unavoidable if progress is to be made. Incidentally, I would include in 'progress' the discovery (if such it were), that control methods, or any other particular approach, were of no use. More to the point, even the most elementary control analysis shows that some correct policies can look rather peculiar when compared with those we are used to. Nonetheless, they may still be worth considering.

May I go on to say that I would take the point of objectivity a good deal further. It is a source of concern to me that whereas we have a number of bodies engaged in quantitative forecasting and policy analysis, we have no objective institute which investigates these matters without a particular axe to grind. It is no disrespect to the groups that currently exist to say that each possesses a fairly high degree of commitment to a particular set of economic doctrines and methods.

What we do not have is a body which contains within it the full range of economic expertise, and is able to apply it to any theory or policy initiative which comes forward. There is a sense in which this is supposed to follow from the debates which occur continously within the economics profession, but in practice the outcome of these, at least nowadays, is confusion.

I would like to see a research institute which is not demonstrably monetarist, Keynesian, interventionist, or free market dominated; which is able to consider econometric methods, con-

trol methods and any other methods dispassionately; which is a repository of data and of established results, and which is a source of sound professional advice to all who ask for it. To use another analogy, in medicine there are institutes which test the efficacy of drugs, searching for dangerous and otherwise unforeseen side effects. More generally, in the sciences there are institutes which concern themselves with the purity of substances and materials. I think more emphasis on efficacy and purity and dangerous side effects in economics would only be for the good.

In this connection may I raise one additional matter which is peripheral to our main theme, but is still important. The objective of much of the work we are talking about is to improve public policy making. In recent years much has been revealed about the UK Treasury's forecasting methods and model, and we are now even given at regular intervals a glimpse of the Tresury's forecasts of some key economic variables. I, for one, assumed that the so-called 'Bray amendment'[2] committed the government to publishing a good deal more than they do. My own opinion is that if the Treasury's forecasts are to be subject to appropriate scrutiny, and benefit from expert criticism, they must all be published. As it is, they are not only kept from outsiders, but most people in the Civil Service are prevented from seeing them as well. This is an absurd state of affairs which must make outsiders fear the worst.

ENGINEERING CONTROL VS ECONOMIC CONTROL

This leads me to two topics which indicate the ways in which engineering control methods are *not* fully relevant to economic forecasting and control. The first of these arises from the fact that the controller of the system is himself part of the system, and that there exist other individuals and groups within the system who are investigating his behaviour. Economists have talked about 'rational expectations', but the issue is more general and more complicated than that. To start with the standard engineering control problem, a rocket going to the moon does not itself forecast the behaviour of the men at NASA, nor does it attempt to duplicate their forecasting and control methods. What is more the controls that are employed on it work directly.

In the economic system the effects of controls may be partly direct, i.e. consumers with lower real incomes spend less. But they are also indirect, for example, business investment depends on business forecasts of future market conditions which in turn depend on their views of what the Treasury think future market conditions will be and what the government is likely to do about them. In other words, business behaviour will be a function of the government's economic objectives, and of estimates of what the government is thought to be forecasting about the future.

Thus, the private investment response to an increase in current sales may itself be an increasing function of the average degree of capacity utilisation which the Treasury is thought to be seeking to maintain. It may also depend on the trade-off assumed to be in the Treasury mind between inflation and unemployment. What we have to recognise, therefore, is that the assumptions according to which the economy is being controlled, or thought to be being controlled, may actually influence the way the economy works and responds to control.

These ideas are essentially game theoretic in character. On the basis of them it is possible to construct models in which the exchange of information, and especially the publication of information by the Treasury, is an essential condition of the system being stable. But it is also easily possible to construct cases where the system becomes unstable or less stable. It is also easy to show that the system may become more stable in the sense of settling down at a low level equilibrium or quasi-equilibrium, but is then extremely difficult to shift to a high level equilibrium.

I may add, parenthetically, that it is especially important to emphasise ideas such as these when teaching economics to engineers and scientists. They will quickly see the connection between much of economics, micro- and macro-, and optimisation theory, and that should make the subject much easier for them at the technical level. It is vital, however, for them also to understand that the optimisation and control of human (or social) systems, es-

[2] An amendment to the Finance Act requiring disclosure by the Treasury of its economic forecasts and the structure of its economic model.

pecially when they are made up themselves of optimisers and controllers, can differ significantly from that of physical systems. That is the reason why management of a firm or of the whole economy is not simply an exercise in mathematics.

My conclusion is that all this is important for the understanding of the UK at the moment. Given what I have just said, we are then led to a second limitation of control theory, namely that many of our problems are those to do with fundamental economic theory, and the shortcomings thereof, rather than simply a failure to apply well established theory. The control engineer can take the established laws of physics and engineering science as given. His problems of estimation of the state of the system or its behavioural structure, are not really tests of scientific hypotheses at the most basic level. In economics the position is different. To take a topical example, the problem of stabilising prices is not solely or even chiefly one of model building, or estimation, or inferring the optimal values of instruments. It is to come to an appreciation, firstly, of how the average level and rate of change of prices is determined, and, secondly, what instruments, if any, influence the various price indices.

We have only an imperfect understanding of how the system works, and we are uncertain about the connection between targets and instruments. Our problem is not what to do in the sense of asking what instruments to apply or what is the best attainable level of the targets. It is what to do in the sense of discovering whether there are any instruments available that might do anything valuable at all to the targets without creating undue damage elsewhere. To use a medical analogy again, we are not choosing among a series of treatments, all of which have some risks, but are pretty sure to work on most occasions. Rather we are looking for any treatment at all that has a serious chance of working without killing the patient. (This last condition is worth stressing. Many economic nostrums of the present day seem to come under the heading, 'The operation was successful, but the patient died.' This is especially true of what might be called the pure monetarist approach to policy.)

May I say finally that I do not wish to exaggerate the difficulties of macroeconomic policy, and it is certainly not my intention to undermine the basic case for more work to be done in the control field. I am rather disappointed that the work in the past 3 years, especially since I ceased to be involved, has not yielded the kind of practical results I had anticipated. It has tended to be too theoretical and methodological. Nonetheless, I think we need to continue the work for several years more, preferably with a high degree of Treasury involvement. We will then be in a position to decide more confidently on its long-term value.

ADDRESS FOR CORRESPONDENCE: *Professor Maurice H Peston, Dept. of Economics, Queen Mary College, University of London, Mile End Road, London E1 4NS, UK.*

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 123–132
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0123/$02.00/0

Strategic Competitive Information Systems

WILLIAM R KING

BIPLAB K DUTTA

JAIME T RODRIGUEZ

University of Pittsburgh

(*Received May 1977; in revised form October 1977*)

A variety of strategic competitive information systems is described and evaluated in terms of three criteria which should be met by strategic information systems. The more sophisticated varieties—those that incorporate specific models into the system—generally meet the criteria better than do the 'retrieval' varieties of systems. Since systems of this kind are relatively expensive to develop, this suggests that the area of strategic competitive information may be one of great potential payoff for the integration of models into information systems.

COMPETITIVE analysis has always been an important element of planning. However, until recently most competitive analysis was performed on an *ad hoc* 'special study' basis by business firms. With the advent of advanced computers and modelling technologies, the capability for developing computerized systems for competitive analysis is becoming more widespread and better recognized.

No person or organization can effectively operate in any competitive environment without a basic understanding of that environment or without up-to-date information on happenings in the environment. This point really needs no extensive logical justification. However, on a pragmatic level, one can demonstrate the high degree of importance being placed on competitive information in planning through the development of intelligence systems by business firms as well as through the growth of more general varieties of 'environmental scanning' as depicted by the increased appearance of departments to perform this function on the organization charts of business firms [6].

The roles which may be played by competitive information in strategic planning are shown in Table 1 in the context of some other varieties of environmental information. This table shows a strategic planning process which is defined in terms of four basic phases—situation assessment, goal development, constraint identification and strategy selection. Table 1 also shows a number of key sources of strategic environmental information in the rightmost column. The middle column depicts a number of information subsystems which may be developed to facilitate environmental scanning.

The various subsystems in the middle column of Table 1 are reasonably self explanatory. Some are non-traditional in nature, like the 'image' subsystem which entails survey data on the perceptions which the organization's clientel have toward it—i.e. the 'image' which it conveys to the environment. Other subsystems—such as the 'customer subsystem'—are well developed in most organizations [9].

Information about competitors appears in three places in the table—once in terms of constraint identification and in two different sub-

TABLE 1. INFORMATION SUBSYSTEMS RELATING INFORMATION SOURCES TO THE PLANNING PROCESS [9]

Strategic planning process	Environmental information subsystems	Strategic environmental information sources
Situation assessment (What is our current situation?)	Image subsystem Customer subsystem	Customers External—e.g. technological, economic and social
Goal development (What do we want our future situation to be?)	Potential customer subsystem Forecasting subsystem	Potential customers External—e.g. technological, economic and social
Constraint identification (What constraints might inhibit us?)	Competitive profile subsystem Regulatory subsystem	Competitors Government
Selection of strategies (What actions should we take to achieve our goals?)	Intelligence subsystem Cost-benefit subsystem	Competitors Other external sources

systems that are related to strategy selection. Competitive information is separated into these elements primarily in terms of *the use which is made of it* rather than the basic nature of the information itself.

In its constraint identification role, competitive information is maintained in a 'competitive profile subsystem'. This subsystem incorporates a competitive profile of each competitor. Such profiles may, in general, be thought of as delineating the business 'character' of the competitor. For instance, one company constructed profiles of all competitors to focus on such factors as:

- background of key competitor personnel;
- characteristics of projects on which competitive proposals were made;
- characteristics of projects on which competitive proposals were not made;
- mix of competitor's in-house business;
- assessment of relative value placed by competition on various performance measures—for example, product quality, service capability, and the like.

In its strategy selection role, competitive information must be incorporated into two information subsystems—an 'intelligence subsystem', which deals with obtaining, or inferring, the answer to specific strategy-oriented questions about competitors, and a 'cost-benefit subsystem' in which competitive information is integrated with other information in a model which is designed to support strategy selection.

In the case of the 'intelligence subsystem', competitive information is of a specific and often, *ad hoc*, nature: what is the competitor's technical strategy likely to be on his next government proposal? Is he likely to begin to compete on a price basis in a particular market? Some of this information is derivable from the profiles which are maintained in the 'competitive profile subsystem', but often, it may be acquired through individual effort directed toward the gathering of data which is related to a specific and timely question about a competitor's likely intentions [1, 3, 12].

In the 'cost-benefit subsystem', information about competition is input into a strategy model which projects the likely consequences of proposed strategies in the context of the market and other environmental factors as well as competitive reactions.

In this paper it is our purpose to describe competitive information subsystems of the 'competitive profile' and 'cost-benefit' varieties which are in operation or under development. 'Intelligence subsystems' are generally much less systematized than these varieties and are treated elsewhere [4]. In dealing with these two varieties of systems, we seek to demonstrate the feasibility of such systems, to illustrate some innovative system concepts, and to evaluate their further potential.

CRITERIA FOR STRATEGIC COMPETITIVE INFORMATION

In order for strategic competitive information to be truly useful in strategic planning,

Vol. 6
1978

it must be evaluated in terms of a number of criteria. These criteria serve to ensure its relevance and usefulness. Three such criteria, which have been developed and applied in a number of firms who have successfully developed such systems, are proposed.

The first criterion for strategic competitive information involves the recognition that much of it will emanate from new sources—often sources that are external to the firm. This implies that *competitive information must be collected and processed in ways that render it into a form which is suitable for the support of strategic choice.* For instance, volumes of newspapers describing competitors' activities or stacks of competitive annual reports are not suitable for direct use in the planning process, except as they may be occasionally used to answer specific questions which may arise. As far as guidance for strategy formulation, data in such raw form are almost useless.

The second criterion for strategic relevance is inherent in the difference between data and information. *Information is evaluated data.* Of particular concern in strategic planning is that *the data be evaluated for some specific strategic purpose or use.* Too often, the information which is developed for the support of strategic planning has been so unfocused as to be of little use in the planning process. Illustrative of this is the competitive data base developed by a large business equipment firm. This 'system' compiled and disseminated one paragraph 'intelligence' statements which were input to it by sales and technical personnel and stored in a computerized memory. These intelligence assessments were prepared after visits to client firms and contacts with competitive personnel. The items in the data base were retrievable only through a 'dump' of all items related to a particular competitor or a specific time period, or both. A systems audit, showed that the system was not utilized to the degree that had been anticipated in the system proposal, that it was not integrated into any decision-making process, and that the lack of evaluation and assessment of items which were input into the system led to a great likelihood of erroneous conclusions being drawn by managers on the basis of the information supplied by the system. As a result of the audit, the system was promptly deactivated.

The third criterion which determines the relevance of strategic information is *the degree to which the information is collected, analyzed and disseminated in a fashion that is integrated with the decision-making processes of the organization.* This will most often mean that the use of information will be integrated into the firm's annual planning cycle, so that it can be used in identifiable ways to influence strategy choice. However, strategic competitive information is also important to the process of strategy implementation and control—i.e. ensuring that plans are carried out and that they are producing the desired effect.

Many companies have developed strategic information 'systems' that have failed because they do not meet this criterion. Often, this is done in the framework of a 'data bank' mentality which says, "Let's collect all of the data that we can identify (concerning competitors, or technology, or whatever area is currently faddish), and it will obviously be of great use to us in formulating our plans". The result is most often a 'room full of data' that is so voluminous, unmanageable, and unedifying that it turns out to have little real value. Many firms have shown, to their dismay, that information systems and processes which do not rate will in terms of these three criteria are probably not worth the effort which goes into them.

In this paper we shall illustrate a variety of operational strategic competitive information processes and evaluate them in terms of these criteria.

COMPETITIVE PROFILE SYSTEMS

There are a variety of operational systems of the general 'competitive profile' variety. These systems generally serve to maintain a profile—in one form or another—of the competitor, his activities or references to him in the public media.

Pure retrieval competitive profile systems

Public sources. Pure retrieval systems are those which have an organized data base of competitive information along with a capability for retrieving information from the data base. Information in the data base is usually based on the analysis of publically-available sources such as magazines, journals and reports of government regulatory agencies. The degree to which such systems meet the previously-

TABLE 2. SAMPLE OUTPUT-PURE RETRIEVAL COMPETITIVE PROFILE SYSTEM-PRIVATE SOURCES

Date	Reporter Origin Products	Agency	Abstract
23 October 1976	FS Smith MD V	C0206-XYZ Corp Q Division Baltimore, MD	The following obtained from an article in local press (1) head of division is GH Jones (2) HB Smith in charge of test. (3) sys at Jones arpt cost about $4.5M (4) Product weighs 12k lbs empty
25 October 1976	PR Johnson PGH W	C0183 SH&F Co. Consultants Baltimore, MD C0206 XYZ Corp Q Division Baltimore, MD	SH&F is a Corp. of planners, architects, and engineers dealing with XYZ type business. They appear to be well structured and employ 150–200 people, two computers, command good fees, and handle cont in 200–300k range. They are working with XYZ on two projects at A and B. XYZ engineers are plant oriented and inexperienced in political environment of this field
29 October 1976	JE Mullins WSH V	C0111 ABD Co. Washington, DC C0206 XYZ Corp Q division Baltimore, MD	Walter of ABD has been calling on XYZ for 10 years. Says they are discontinuing negotiations on project A and that all work has stopped in dispute on project B.

established criteria is generally minimal, except that systems which are supported by a good analysis function—which assesses what should go into the data base and in what form—can meet the former two criteria to some degree.

Private sources. Pure retrival systems may also be developed to collect and disseminate personal reports of items of interest about competitors. These systems are different from those which are based on publically-available data in that they encompass private reports such as those made by sales personnel who may be charged with reporting on contacts that they make with customers, competitors, and others in which facts or rumors about competitors are brought to attention.

A typical response to an inquiry into such a system is shown as Table 2. Brief reports made by field personnel may be retrieved by inquiring about a particular competitor, reporter, or time period. The system produces all reports for the category or combination of categories requested. Such systems have been referred to earlier in a negative context. The authors know of no automated systems of this variety which meet the previously-noted criteria and which have been evaluated to be sufficiently useful to their user-managers to be cost-effective.

Abstract retrieval systems

Another variety of retrieval system which has been developed by some firms is similar to a library retrieval system. For instance, one company has a system which provides managers with access to a data base of publically-available document abstracts. These abstracts describe articles which have appeared concerning specified competitors in more than 500 sources—technical journals, business publications, press releases, Securities and Exchange Commission filings, and newspapers.

The data base now contains over 10,000 abstracts which have been developed by a staff who scan these publications and other documents. The staff people write a brief of each article and index the article in terms of: the competitive company to which it relates; the relevant product line(s); the topic content of the article. Each of the descriptors in the three categories is chosen from a list of those with which the system is capable of dealing. For in-

TABLE 3. SAMPLE PARTIAL OUTPUT-ABSTRACT RETRIEVAL SYSTEM

```
DATA BASE NAME: COMPETITIVE ABSTRACT
DATA BASE UPDATED—JUNE 21, 1977
ENTER DESCRIPTOR: XYZ COMPANY
482 ARTICLES OUT OF 10,236
ENTER ANOTHER DESCRIPTOR: WIDGETS
36 ARTICLES OUT OF 482
ENTER ANOTHER DESCRIPTOR: PRINT
ARTICLE 4237
SOURCE: BUSINESS WEEK, 6, 20, 77, p 48
XYZ COMPANY ANNOUNCED PURCHASE OF
WIDGET COMPANY OF AMERICA IN DEAL THAT HAD BEEN RUMORED FOR
MONTHS. XYZ PAID 20 PERCENT ABOVE MARKET PRICE FOR 80 PERCENT
OF OUTSTANDING SHARES. MARKET PRICE OF WCA STOCK HAD PREVIOUSLY
DOUBLED THIS YEAR. FINANCIAL ANALYSIS QUESTION PRICE PAID.
RUMORS OF QUICK EXPANSION PLANNED FOR WCA HAVE CIRCULATED
SINCE PURCHASE.
                    :
```

stance, the topical content list contains descriptors such as 'patent', 'legal', 'financial', and 'mergers and acquisitions'. Users of the system may inquire in terms of any of the indexing categories. One might for example, specify a particular competitive company, or a specific product line, or a topic area. The system responds to each inquiry with an indication of the number of abstracts in the data base to give the user some idea of the magnitude of the search on which he is embarking.

Normally, a user successively specifies more than one index category. A typical inquiry is depicted in Table 3. There, the user begins using the system by specifying the 'Competitive Abstract' data base. The system responds by indicating when the data base has most recently been updated, and asking the user to enter a descriptor. In this case, he first enters a competitive company name, although he could have used either a topic descriptor or a product line descriptor.

The system responds by identifying the number of articles in the data base which are indexed as relating to this company and asks for another descriptor. In this case, the user then enters a product line. In effect, he has said that he wishes to know only about XYZ's widget line at this point. Again he is given the number of articles in this dual category and the system asks him to enter another descriptor. Instead, he types in 'PRINT', indicating that he does not wish to further delineate the category, but instead desires a listing. The system such as this is much like those which some libraries use to retrieve 'information about information'—i.e. to retrieve secondary information concerning the source of primary information. In this case, the primary information is also conveyed in abstract form, whereas in other instances, only the reference is incorporated into the data base. If the user desires to have the primary information—in this case the *Business Week* issue in which article 4237 appeared—he must go the company library where it is filed by the article number.

While such systems may be generally useful as sources of information about competitors, they do not meet the criteria discussed earlier since they provide only partially-evaluated data and they are not integrated into the planning and decision making processes of the company.

COST-BENEFIT COMPETITIVE SUBSYSTEMS

The variety of planning subsystem which is encompassed by the 'cost-benefit' category is that which produces more significant compilation of information or projections of the likely consequences of proposed strategies. In effect, such systems produce data which have been 'evaluated' in a very sophisticated manner to produce information.

If this evaluation is performed by using competitive data as an input to a 'corporate model', the competitive data may be used to develop projections which can serve as the model's answer to 'What if?' questions as posed by its users.

While such systems are well developed in many firms, their competitive modules do not appear to be nearly as sophisticated as many other aspects of the overall corporate milieu [10]. This is probably due to the inherent difficulty in modelling competitive situations. Here, we shall treat two innovative approaches to strategic competitive modelling which are quite different than those that have been previously developed for the support of strategic planning. The first—a 'strategic issue' approach—is really a hybrid version of a 'competitive profile' system and a 'cost-benefit' system. The second, which uses meta-game analysis of strategies, is a true member of the 'cost-benefit' category.

STRATEGIC ISSUE COMPETITIVE INFORMATION SYSTEMS

A strategic issue competitive information system (SICIS) is one that is developed around 'strategic issues' which face the planners in an organization. One such system which has been developed and tested will be described here. Another system, developed on the basis of 'strategic questions' has been reported by Naumes.[1]

The SICIS system is one which is designed to provide strategically-useful information about competitors without requiring the user to know exactly which elements of data may be required. The basic concept is that the user should be able to make inquiries of the system in terms of the *strategic issues* with which he is concerned. The SICIS system can be best understood by contrasting it with the 'pure retrieval' systems. As noted earlier, a 'pure retrieval' system is one which permits the user to request any item or combination of items in the data base, so long as he knows in advance eactly *what* he wants, and so long as he understands the operation of the system sufficiently well to know *how* to request what he wants. Strategic managerial needs for information are not so clean cut; thus, such systems seldom measure up to the critical critera for strategic information.

A 'strategic question' competitive information system has a data base which is structured in terms of *strategic business issues.* This permits inquiries to be made in terms of *the use to which data are to be put, rather than in terms of the data themselves.* Thus, a manager using such a system inquires with a question such as, "What is the capability of Competitor X to introduce a new product in his Line Y next year?", and is presented with system responses which successively identify subclasses of data in which he may be interested. For instance, such a query might produce a system response that would identify Competitor X's financial capability, production capability, marketing capability, and technological capability, as elements of the overall issue.

The user could then identify specific areas of interest, or he could indicate that he desires the total picture related to his initial question. A user indication of interest in marketing capability would produce a system response indicating the availability of data on competitive distribution channel capacity, field sales capability, service capability, technical sales expertise and a variety of other marketing-related areas. The system would also indicate its ability to provide projections of future market growth. Indications of interest in other areas would produce similar system responses which would, after each new user response, indicate successively more-detailed sets of available data.

Table 4 shows the print-out from one interactive inquiry into the SICIS system. It shows how the system prompts the user by specifying the sub-issues which make up the information structure model underlying the issue into which he has inquired. This permits him to refine his query if he really does not desire information for as broad a spectrum as the initial inquiry specifies. It also serves to suggest related areas to the user that may go beyond his personal conceptualization of the issue which is the basis for his inquiry. Of course, such a system is limited to strategic issues and questions which have been preprogrammed. However, if the system is designed, as is the one being described here, on the basis of those that are most urgently needed in supporting their strategic choices, the system is likely to be used. A system such as SICIS meets the first two of the previously-outlined criteria at least as well as do other retrieval-oriented systems. In terms of the third criterion, its basic

[1] NAUMES WILLIAM, *Computer Aided Decision Making*, unpublished working paper, Temple University, 1977.

TABLE 4. STRATEGIC ISSUE COMPETITIVE INFORMATION SYSTEM

ENTER CODES OF WHICH COMPETITORS TO EXAMINE
01—TAPPAN CO.
02—DESIGN AND MANUFACTURING
03—SEARS CO.
04—GENERAL ELECTRIC CO.
99—ALL COMPETITORS
1,4
ENTER ISSUE CODE OF DESIRED STRATEGIC ISSUE
IF NO MORE ISSUES DESIRED ENTER 00
2
2—WHAT RESOURCES LIMIT THE COMPANY'S/COMPETITOR'S ABILITY TO ACHIEVE GOALS?
1—FINANCIAL CONDITIONS.
2—MANAGEMENT CONSTRAINTS.
3—MANUFACTURING EFFICIENCY.
4—MARKETING OR DISTRIBUTION SYSTEM.
5—RESEARCH AND DEVELOPMENT.
6—IMPOSED RESTRAINTS.
ENTER CODES OF DESIRED COMPONENTS IF ALL ENTER CODE 99
1,4
1—FINANCIAL CONDITIONS.
(1) PRESENT LOAN LEVELS.
(2) QUANTITY OF FUNDS THAT CAN BE FINANCED INTERNALLY.
(3) CAPITAL STRUCTURE.
ENTER CODES OF DESIRED COMPONENTS IF ALL ENTER CODE 99
2
2—QUANTITY OF FUNDS THAT CAN BE FINANCED INTERNALLY.
TAPPAN CO.
(MILLIONS)

	1970	1969	1968	1967
SALES	313	133.8	124.2	95.7
PROFITS	2.23	3.44	3.97	1.00
RETURN ON ASSETS	2.8%	4.5%	5.3%	1.9%
% PROFIT/SALES	1.7%	2.6%	2.6%	1.0%
TEN YEAR ROA	5.9%			
TEN YEAR PROFIT/SALE		2.8%		

GENERAL ELECTRIC CO.
(MILLIONS)

	1970	1969	1968	1967
SALES	8.726	8.448	8.382	7.741
PROFIT/SALES	3.8%	3.3%	4.3%	4.7%
TEN YEAR ROA	17.2%			
TEN YEAR PROFIT/SALES		4.6%		

D—MARKETING OR DISTRIBUTION.
(1) OWN VS. INDEPENDENT DISTRIBUTION.
(2) AGGRESSIVE OR PASSIVE MARKET ORGANIZATION.
ENTER CODES OF DESIRED COMPONENTS IF ALL ENTER CODE 99
1
(1) OWN VS. INDEPENDENT DISTRIBUTION.
TAPPAN CO.
TAPPAN COUNTED WITH ABOUT 120 SALESMEN WHICH SOLD TO RETAILERS, LARGE BUILDERS, AND INDEPENDENT BUILDER DISTRIBUTORS. TAPPAN WAS CHANGING ITS DISTRIBUTION SYSTEM AND WAS SETTING UP ITS OWN CENTRALIZED AND REGIONAL WAREHOUSES.
TAPPAN FRANCHISED ITS SERVICING CAPABILITY TO PRIVATE COMPANIES.
GENERAL ELECTRIC CO.
GE CONTROLS ITS OWN CHANNELS OF DISTRIBUTION AS WELL AS SERVICE OUTLETS AND HAS A POTENTIAL ADVANTAGE IN SURROUNDING SEARS WITH COVERAGE USING THE GE BRAND, HOTPOINT AND PENNCREST BRAND.
ENTER CODES OF WHICH COMPETITORS TO EXAMINE
01—TAPPAN CO.
02—DESIGN AND MANUFACTURING
03—SEARS CO.
04—GENERAL ELECTRIC CO.
99—ALL COMPETITORS
ENTER ISSUE CODES OF DESIRED STRATEGIC ISSUE
IF NO MORE ISSUES DESIRED ENTER 00
00
THANK YOU FOR USING SICIS TRY US AGAIN
EXIT

structure—that of strategic issues—enables it to be better integrated into the planning process than can most such systems.

COMPETITIVE SCENARIO MODELLING SYSTEMS

The most sophisticated level of competitive information systems is that which clearly fall into the general 'cost-benefit' subsystem category. Such systems attempt to evaluate the costs and benefits associated with proposed strategies to permit the choice of a 'best', or a 'good' strategy. Any corporate model which entails competitive dimensions is such a system. However, the competitive aspects of corporate models are often their weakest elements. One such system which emphasizes the competitive aspects of strategy is currently under development in cooperation with a major company. The system, called COSMOS (Competitive Scenario Modelling System) involves 'competitive scenarios' as a basis for utilization of the system. A *competitive scenario* can be visualized as a particular *set of strategy choices* involving price, product quality and other controllable elements which are made by a firm *and* its competitors in a market.

A manager may think of a competitive scenario and ask himself the question, "How credible is this particular scenario when I look at it from *both* the points of view of my firm and that of my competitor?" A scenario can be considered credible (likely to come about, or 'stable') if each competitor (including the firm) thinks it is desirable on its part to stick to its strategy choice given in the particular scenario, after considering alternative moves on its part and possible counterreactions from others.

Using COSMOS interactively, a manager can systematically explore various scenarios and by comparing them with the test scenario can decide whether there are any strategy choices on his part which can lead to an improved position for his firm over the test scenario. The system also permits consideration of possible countermoves on the part of competitors that might block any attempt at improving the firm's position over the test scenario.

The COSMOS system design integrates Howard's analysis of options technique [7] which is based on the theory of metagames [8, 11] with mathematical market models and multi-attribute decision making. Figure 1 shows that after specifying the test scenario to be examined for stability, a firm (i) is chosen from whose viewpoint the first run of the exploration will be made. Assuming initially that the other competitors will keep to their part in the scenario, alternative scenarios are generated by varying the strategy choice of i while keeping the strategies of others unchanged. Each alternative scenario is evaluated in terms of its market performance (market share and profits) for firm i, and compared to the performance implied by the test scenario.

If no preferred alternative scenario can be found, that is, if no *unilateral improvement* can be discovered, the test scenario is the best position that can be achieved by i, under the condition that others stick to their choices in the test scenario. Then, this procedure is repeated for all the other competitors and if no unilateral improvement can be found for any party, the test scenario can be considered stable for each competitor.

If one or more unilateral improvements are found for firm i, the system proceeds to the next step of considering countermoves on the part of other parties than can effectively block any move by i to improve his position. Any such countermoves are called *sanctions*. To find any possible sanctions, the strategy choices of the other parties are varied and for each combination of their choices, a search is made for a reply, that is, a strategy choice by i such that the scenario generated is preferred to the test scenario. If no such reply can be found, the specific combination of other parties' strategy choices constitute a *sanction*. If some sanction is found and there is no reason to doubt its credibility, the position is the test scenario and is the best that i can hope for. Then, the system returns to step 3 to consider another competitor from whose viewpoint further testing will be done. If no sanctions exist against firm i, the system proceeds to looking for all guaranteed improvements for i from the test scenario. A guaranteed improvement can be found when we obtain a strategy choice such that whatever be the strategy choices of other parties, the resulting scenario is preferred by firm i to the test scenario.

The analysis procedure can be repeated for

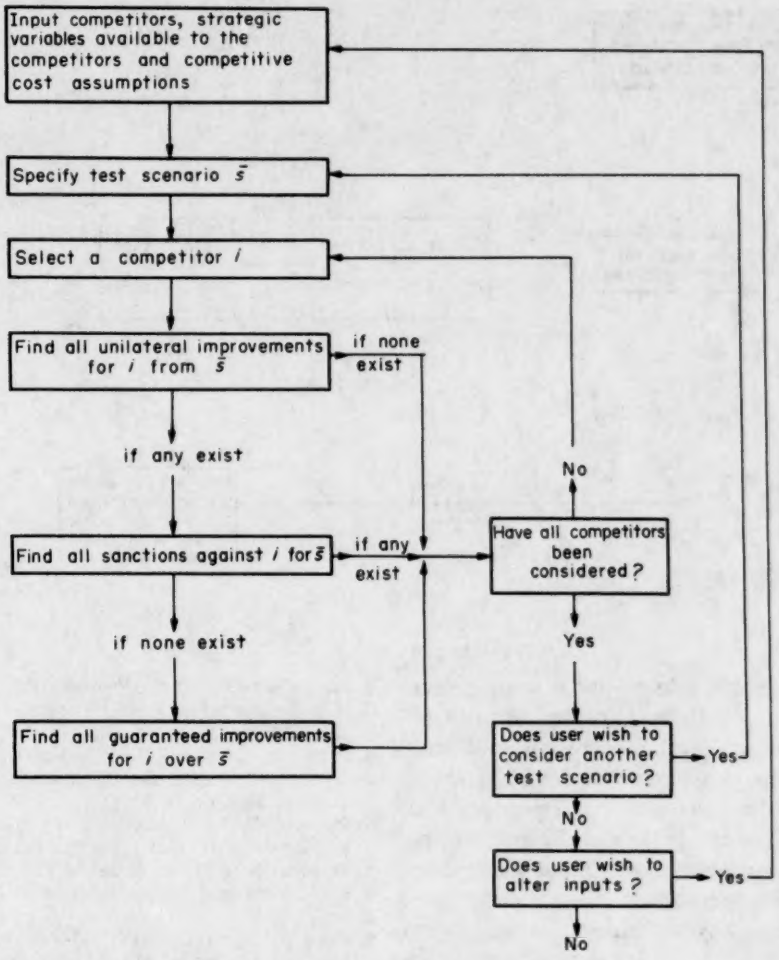

Input competitors, strategic variables available to the competitors and competitive cost assumptions
Specify test scenario $\bar{s}$
Select a competitor i
Find all unilateral improvements for i from $\bar{s}$
if none exist
if any exist
Find all sanctions against i for $\bar{s}$
if any exist
if none exist
Find all guaranteed improvements for i over $\bar{s}$
No
Have all competitors been considered?
Yes
Does user wish to consider another test scenario?
Yes
No
Does user wish to alter inputs?
Yes
No

FIG. 1. *Flow chart for competitive scenario exploration.*

any other scenario to be tested, and for a modified set of available strategies as well as for modified assumptions concerning cost and other situational parameters to explore the decision-situation in depth. The potential for modifying underlying assumptions allows for sensitivity analysis as well as for the testing of scenarios under conditions where the various competitors' understanding of the market is different. This permits consideration to be given to the situation in which one firm's market model is different from that of a competitor—i.e. a situation in which the players see themselves as playing in different games.

Figure 2 shows a schematic diagram of how the overall COSMOS system operates. The input to the system is a list of competitors, a list of possible strategies—each of which is a set of specified values for a set of strategic variables, such as price and quality—and the identification of a test scenario to be initially evaluated. Each scenario which is generated from the previously-described exploration process is evaluated for its market performance using market share and profit models. These models utilize the judgemental inputs of user-managers. The 'preference-aiding' model shown in Fig. 2 is one which records the prior choices made by user-managers in choosing among outcomes described in terms of profit and market share, so that the user is able to see his pattern of prior preferences in expressing his preferences for such outcomes.

A cost-benefit system such as COSMOS serves to well fulfill the criteria for strategic competitive information since it directly

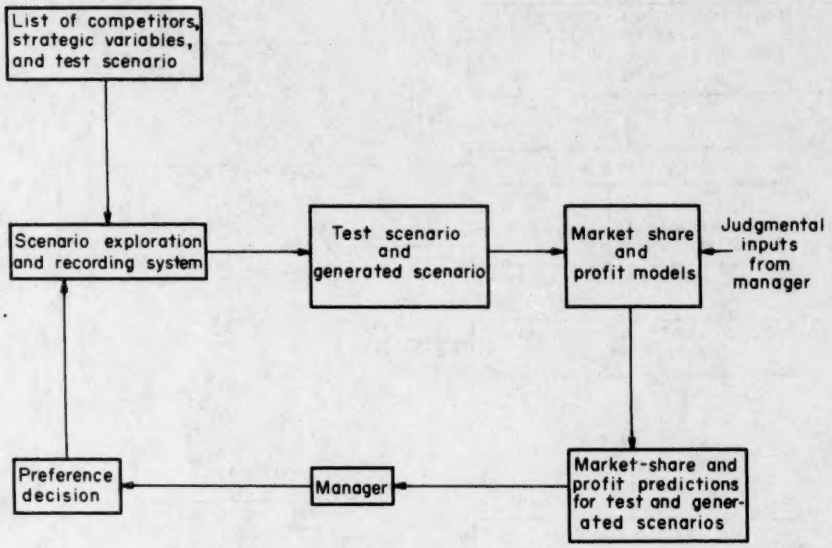

List of competitors, strategic variables, and test scenario
Scenario exploration and recording system
Test scenario and generated scenario
Market share and profit models
Judgmental inputs from manager
Preference decision
Manager
Market-share and profit predictions for test and generated scenarios

FIG. 2. *Overall system.*

focuses competitive information in specific decision-oriented directions. However, as with all such systems, its greater degree of specificity means that more assumptions are being made concerning the behavior of competitors and competitors' models of the world. However, the varying of assumptions and input parameters can vitiate this weakness if they are done comprehensively and systematically.

REFERENCES

1. BARCLAY S (1973) *The Secrets Business.* Thomas Y Crowell, New York.
2. *Business Week* (August 4, 1975) Business sharpens its spying techniques.
3. CHEMICAL ENGINEERING (1966) *Gathering Competitive Information.*
4. CLELAND DI & KING WR (1975) Competitive business intelligence. *Bus Horizons* **18**(6), 19–28.
5. DE CARBONNEL FE & DORRANCE RG (1973) Information sources for planning decisions. *California Mgmt* **XV**(4), 42–53.
6. FAHEY L & KING WR (1977) Assessing the impact of environmental scanning in corporate planning. *Bus. Horizons* **20**(4), 61–71.
7. HOWARD N (1975) Analysis of options in business problems. *INFOR* (Can. Jl Op. Res. Inf. Sci.) **13**, 48–67.
8. HOWARD N (1971) *Paradoxes of Rationality: Theory of Metagames and Political Behavior.* MIT Press, Cambridge, MA.
9. KING WR & CLELAND DI (1974) Environmental information systems for strategic marketing planning. *J. Marketing* **38**(4), 35–40.
10. NAYLOR TH & SCHAULAND H (1976) A survey of users of corporate planning models. *Mgmt Sci.* **22**(9), 927–937.
11. RADFORD KJ (1975) Application of metagame theory in managerial decision making. *Omega* **3**(3), 303–312.
12. WALL JL (1974) What competition is doing: your need to know. *Harv. Bus. Rev.* **52**(6).

ADDRESS FOR CORRESPONDENCE: *Professor William R King, Graduate School of Business, University of Pittsburgh, Pittsburgh, PA* 15260, *USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 133–138
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0133/$02.00/0

Company Structure and Market Strategy

R MANSFIELD

University of Wales Institute for Science and Technology

D TODD

J WHEELER

Imperial College, London

(*Received September 1977; in revised form October 1977*)

The importance of the relationships between company strategies and structures is noted, as is the relative dearth of empirical work in this area. The article reports a preliminary study of 24 companies investigating the relationships between strategies of diversification and geographic dispersion of territories covered and structural variables relating to the number of structural differentiations in a company, the decentalization of decision-making and functional specialization. Contrary to the hypotheses proposed, these two strategic variables have different structural implications with geographic dispersion particularly associated with structural differentiation and diversification associated with decentralization.

RECENT writers in the field of organization theory have examined the interactions between organizational environments and structures, (see [16, 5, 20, 1]). They have also investigated the implications for organizational performance of the 'fit' between structure and environment. However, they have often ignored the concept of organizational strategy in formulating empirical research designs and theoretical arguments. It has been left primarily to business historians, following the seminal work of Chandler [6] to disentangle the interrelationships between company strategies and environmental and structural variables (e.g. [7, 11]). Chandler [6, p. 15] summarized the interrelationships he found as follows: "strategic growth resulted from an awareness of the opportunities and needs created by changing population, income and technology —to employ existing or expanding resources more profitably. A new strategy required a new or at least refashioned structure if the enlarged enterprise was to be operated efficiently."

Unfortunately the methodologies employed by business historians and organization theorists have diverged to such an extent that it is difficult to draw togeether the research findings from the two areas. This paper reports the findings of the preliminary stage of a study designed to examine the relationships between market factors, company strategies and structures. The present sample consists of 24 companies operating in the construction, food processing, paper and board and machine tool industries.

Business strategy

A company's business strategy is here considered essentially in terms of the market or markets in which it attempts to sell its goods or services. Such a conceptualization fits well with the ideas of Chandler [6], Drucker [9] and Ansoff [3] all of whom see strategy as largely revolving around the choice of product markets and of the territory in which the company is to operate. Such an approach to strategy is more limited than that advocated by some writers (e.g. [15, 17]) but does include the most vital aspects of the concept. Strategy refers to a company's long term policy both for defining and dealing with its environment, and in the capitalist or mixed economy the

market dimension is likely to be the most important. Indeed Chandler [6, p. 383] concluded as a result of his study of the growth of some of the largest American companies that "the common denominator of structure and strategy has been the application of the enterprise's resources to market demand."

The present study therefore concentrates on two aspects of market strategy; first, the extent to which a company provides a diverse range of products or services; and second the geographic dispersion of the territories in which sales are carried out. These measures together define the extent of the markets in which the company has an interest, which is likely to have significant implications for company structure.

The creation of strategies and structures

Clearly company strategies and structures are mainly the consequence of decisions taken by those in power within constraints imposed by both internal and external factors. Ansoff [3] in an attempt to develop a theoretical framework for the study of business strategy has devised a classification of company decision-making in which he distinguishes three types of decisions. First those which define the company's *strategy*. Second those which define its structure in terms of the internal arrangement and administrative procedures which will be used in order to best implement its strategies. Third operating decisions which concern the day-to-day conduct of the enterprise. These decisions fix the short term tactics which detail the operation of the company's production and marketing processes. Ansoff argued that these three decision levels are linked in a cybernetic way, but he clearly believed that strategic decisions are usually the first to be taken in any particular sequence. Chandler argued that such conclusions concur with his findings although consideration must also be given to the growth stage of the organisation. Thus as a company grows attention oscillates between the need to expand and diversify in order to exploit markets and make full use of technology and personnel and the need to rationalise the company's structure and control system.

The interaction between strategy and structure depends upon the environmental conditions in which the company operates. Even here the relationship will reflect patterns of causality in both directions as companies not only react to their environments but seek actively to change them. Many writers have argued that companies regularly try to take the initiative and manipulate the market forming cartels and relationships with governments as well as by advertising (see, e.g. [6, 10, 12, 14, 21]).

Product or market diversification is occasionally achieved by an accidental sequence of events such as an unplanned technological advance [13]. Usually, however, diversification is strategically planned and implemented over a long period of time. Such strategies are often adopted as a response to external situations of risk, dependence or heavy competition [13, 18]. Andrews [2] on the other hand emphasised the role of internal pressures when he suggested that diversification may be aimed either at achieving a more effective use of company resources or providing for a more efficient marketing strategy.

Chandler [6] concluded that in the US strategies of diversification led almost inevitably to the adoption of multidivisional structures. Franko [11] reported that even in diversified companies divisionalization had not occurred to the same extent in Europe up to the late 1950's although it has increased since. He argued that divisionalization is due to the interaction of a strategy of diversification with a competitive market situation. It is to be expected that greater diversification of products or services or indeed greater dispersion of marketing territories will create pressures for divisions within a company. Hence it is hypothesised that the more a company is diversified the greater will be the number of structural differentiations within the company (hypothesis 1a), also the greater the degree of geographic dispersion of the territories covered by a company the greater the number of structural differentiations (hypothesis 1b). Clearly, increasing the diversity of products and services offered by a company will increase the difficulty of top level management in maintaining personal knowledge and involvement in decision-making. It is therefore hypothesised that the greater the level of diversification the greater will be the extent of decentralization of decision-making (hypothesis 2a). Despite the sophistication of modern communications it is still likely that geographic

Vol. 6
1978

dispersion will also inhibit the personal involvement of senior management in decision-making. Hence it is predicted that the greater the extent of geographic dispersion of the territories covered by a company the more decentralized will be its decision-making (hypothesis 2b). It it is true that greater diversification and geographic dispersion are associated with greater structural differentiation and decentralization of decision-making then it is clear that such forces will create the need for other means of control. The alternative to centralized decision-making involves the use of bureaucratic rules, procedures and paperwork as well as the utilization of specialists in staff positions [8]. It is to be expected that greater diversification will be associated with a higher degree of functional specialization (hypothesis 3a); the greater the degree of geographic dispersion the higher will be the extent of functional specialization (hypothesis 3b). In the remainder of the article these hypotheses are examined empirically.

The measures used

Although diversification is seemingly a simple concept, it is not easy to create a meaningful operational measure of it which allows comparison of companies in widely differing industries. Clearly a measure based upon the number of products produced gives no real idea of the diversity of markets in which a company operates. Channon [7] suggests that measures of diversification should be based upon the extent of difference in the technology which is employed to make different products. However, such an approach appears to make too many assumptions about the relationship between internal production processes and external marketing strategies. In the present context a measure based upon the Standard Industrial Classification was utilized. This measure was based upon three lower order measures which assessed (1) the number of main industrial market sectors (two digit figures) in which the company was involved, (2) the number of sub-sectors (four digit figures) in which the company was involved, and (3) a measure inversely related to the percentage of total revenue stemming from any single sub-sector. Clearly there is an element of arbitrariness in such a measure but it should give a relatively reliable indication of the extent of diversification as it is commonly understood. A meaningful operational definition of geographic dispersion is similarly difficult to create. This problem is exacerbated by the potentially different implications of dispersion within one country and dispersion across national boundaries. The measure used here was a combination of six lower-order measures, each indicating an aspect of dispersion at home or abroad. (Precise details of this and other measures may be obtained from the authors.) These six measures were designed to indicate the extent to which the companies sell nationally in the UK rather than regionally and also the extent to which the company operates abroad. In fact the six lower order measures are strongly intercorrelated hence it is possible to have confidence that the aggregate measure genuinely represents important aspects of the concept.

The first of the structural measures indicated the number of different structural differentiations in a company. This measure is based upon a conception of organisations as nested systems of departments, divisions and companies differentiated in most cases either along territorial, product or functional lines. The measure indicates the number of levels of nesting in a company. This should indicate the extent to which the administrative structure of a company is differentiated but does not indicate the locus of decision-making.

The second measure concerned the extent of decentralization of decision-making which was assessed from responses concerning the lowest level in the hierarchy at which particular decisions could be made, measured on an eight point scale ranging from 0 for decisions taken by controlling interests outside the organisation and 1 for decisions taken by the company's board through intermediate levels to a score of 7 for decisions decentralized to operative level. The overall measure was obtained by averaging the extent of decentralization with regard to eight different decisions such as that to go ahead with the development of a new product or fixing product prices. The third structural measure indicated the degree of functional specialization in the company—whether or not one or more full-time specialist was employed in each of eleven different functional areas such as buying and advertising. This can be broken down into two sub-scales,

the first covering six customer related functions, the second concerning five functions not directly customer orientated. Although the structural measures were devised for this project they are based upon the techniques used by the Aston group [19] and bear a close resemblance to some of their measures.

RESULTS AND DISCUSSION

The relationships between the measures of strategy and structure were explored by correlational analysis; the zero order Pearson correlation coefficients being shown in Table 1.

Examination of the results shown in this table indicate that the hypotheses are not uniformly confirmed, and that diversification tends to have rather different structural implications from those of geographic dispersion. Hypothesis 1a is not supported by the data as no significant relationship was found between the extent of diversification of products or services in a company and the number of structural differentiations, although there was a weak relationship in this direction. Hypothesis 1b, however, is very strongly supported by the evidence of this study. Clearly even with modern means of communication geographic dispersion of activities leads managers to create more divisions, sub-divisions and departments within their companies. Without doubt this particular phenomenon is partly due to the different legal requirements of the different countries in which many of the companies studied operated, although examination of the correlations between the sub-scales of the dispersion measure and the number of structural differentiations shows all aspects of dispersion have a similar effect. When the relationships between the measures of strategy and the decentralization of decision-making are examined support is again found for one hypothesis but not the other. Hypothesis 2a is supported by the data from the present study where a moderate positive relationship was found between the extent of diversification in a company and the decentralization of decision-making. No relationship at all was found however between geographic dispersion and the locus of decision-making. These results are no doubt influenced by the type of decision about which information was requested, since the questions related to issues which might well be regarded as representing one area of policy issues. The results suggest therefore, that in this area at least, diversification creates pressure for decentralization whilst geographic dispersion does not.

TABLE 1. THE RELATIONSHIPS FOUND BETWEEN MEASURES OF STRATEGY AND STRUCTURE. ZERO ORDER PEARSON CORRELATION COEFFICIENTS ARE SHOWN IN EACH CASE

	Extent of diversification	Geographic dispersion
Number of structural differentiations	+0.25	+0.80***
Decentralization of decision-making	+0.47**	0.00
Overall functional specialization	+0.25	−0.10
Customer related functional specialization	+0.15	−0.36*
Non customer related functional specialization	+0.34	+0.39*

* Indicates $P < 0.05$.
** Indicates $P < 0.01$.
*** Indicates $P < 0.001$.

With regard to the third set of hypotheses it can be seen from Table 1 that neither hypothesis 3a or 3b is supported as no significant relationships are found between overall functional specialization and either of the measures of strategy (although a correlation of 0.25 is found between overall functional specialization and the extent of diversification). Examination of the relationships between the measures of strategy and the two sub-scales of functional specialization suggested that the insignificant correlations reported above may be hiding some important relationships that do exist. It can be seen that in the case of the five non-customer related functions there is a significant positive relationship between functional specialization and geographic dispersion and a positive relationship with diversification which was not quite significant at this sample size. When attention was turned to the six customer related functions the relationship found with diversification was very low but there was a significant negative relationship with geographic dispersion. Clearly the need to break up the overall set of activities in the sales and marketing area by territory reduces a company's ability or wish to subdivide these activities by function. The weakness of the relationship between specialization and diversification may be due to similar forces resulting from the need to have some sales and marketing activi-

Vol. 6
1978

ties differentiated by product rather than function. An overall view is that diversification and dispersion may require the employment of specialists in some functional areas but that in those areas most immediately concerned with customers the requirement for specialization along territorial or product lines reduces the extent of functional subdivision.

One problem concerning the above results concerns the extent to which the relationships reported are due to the influence of size. Certainly Chandler's study [6] suggests that the relationships between strategy and structure may depend on the growth stage of the company and other research shows that organizational size is associated with increased internal differentiation [4] with decentralization of decision-making and with increasing functional specialization [20, 8]. The results of the present study show an approximately similar pattern with company size (operationalized as the logarithm of the number of employees) being positively related to the number of levels of functional differentiation ($r = +0.25$), decentralization of decision-making ($r = +0.25$), and functional specialization ($r = +0.43$) although only the last of these correlations is significant at the 5% level. However, in this sample company size is unrelated to the extent of diversification ($r = 0.02$) so that the effects of these two variables upon company structure must be independent. However, the case of geographic dispersion is more complex as this is significantly related to company size ($r = +0.47$). The possible effects of size upon the correlations between geographic dispersion of territories covered and structural variables was examined by partialling out the effects of size. When this was done the correlation with the number of levels of functional differentiation remained unchanged but those with the various scales of functional specialization were somewhat altered. The partial correlation with specialization in non-customer related functions was reduced to 0.33 which is just below the 5% level of significance. However that with specialization in customer related functions was increased negatively to -0.57 and that with overall function specialization to -0.38 which is significant at the 5% level. When the effect of company size is controlled in this way it is doubly clear that hypothesis 3b is incorrect and that the general tendency created by geographic dispersion is to reduce functional specialization particularly in small companies, although this effect only occurs in the functions directly related to the customer.

In order to investigate directions of causality information was also obtained on the extent to which companies had diversified into new industrial sectors in recent years and a measure based upon increased diversification in the last 5 and 10 years obtained. The correlations between this measure and the structural measures were computed. Only one significant correlation was obtained—a positive relationship between recent increases in diversification and the formalization of market research activities ($r = 0.51$). This is scarcely a surprising conclusion and is almost certainly a consequence of causal influences running in both directions. With this one exception then the present study provides no evidence that company structures are related to increases in diversification at least in recent economic conditions. This suggests that structures do not cause changes in strategy. This being the case, if the relationships between strategy and structure are confirmed it is tempting to conclude that structure generally follows strategy as Chandler [6] and Ansoff [3] suggest.

CONCLUDING COMMENTS

Despite its small scale, the preliminary study reported here provides evidence that company structures are related to strategy, and at least some grounds for inferring that they are partly caused by it. This suggests that much previous research attempting to relate organizational environments directly to structure ignores important intervening variables which must be considered if our understanding is to improve significantly. Certainly if the conclusions of such research are to be useful to managers the role and importance of strategy must be fully explored.

The relationships reported here must be regarded with caution until they are confirmed (or otherwise) on larger samples but they point to important research possibilities.

REFERENCES

1. Aldrich HE & Pfeffer J (1976) Environments of organizations, in *A. Rev. Sociol.* II, 79–105.

2. ANDREWS KR (1951) Product diversification and the public interest. *Harv. Bus. Rev.* **29**, 91–107.
3. ANSOFF HI (1965) *Corporate Strategy*. McGraw Hill, New York.
4. BLAU PM (1970) A formal theory of differentiation in organizations. *Am. Sociol. Rev.* **35**, 201–218.
5. BURNS T & STALKER GM (1966) *The Management of Innovation*. Tavistock, London.
6. CHANDLER AD (1962) *Strategy and Structure*. MIT Press, Cambridge, MA.
7. CHANNON DF (1973) *The Strategy and Structure of British Industry*. Macmillan, London.
8. CHILD J (1973) Predicting and understanding organization structure. *Admin. Sci. Q.* **18**, 168–185.
9. DRUCKER P (1954) *The Practice of Management*. Harper & Row, New York.
10. EDWARDS CD (1964) *Cartelization in Western Europe*. US Dept of State, Printing Office, Washington DC.
11. FRANKO LG (1974) The move toward a multidivisional structure in European Organizations. *Admin. Sci. Q.* **19**, 493–506.
12. FREEMAN JH (1973) Environment, technology, and the administrative intensity of manufacturing organizations. *Am. Sociol. Rev.* **38**, 750–763.
13. HABER LF (1971) *The Chemical Industry* 1900–1930. Clarendon Press, London.
14. JONES R & MARRIOTT O (1970) *Anatomy of a Merger: A History of G.E.C., A.E.I. and English Electric*. Jonathan Cape, London.
15. KHANDWALLA PN (1976) The techno-economic ecology of corporate strategy. *J. Mgmt Studies* **13**, 62–75.
16. LAWRENCE P & LORSCH J (1967) *Organization and Environment*. Harvard University Press, Cambridge, MA.
17. MOORE D (1959) Management strategies, in *Industrial Man: Businessmen and Business Organizations* (Eds. WARNER WL & MARTIN NH). Harper, New York.
18. PFEFFER J (1972) Merger as a response to organizational interdependence. *Admin. Sci. Q.* **17**, 382–394.
19. PUGH DS, HICKSON DJ, HININGS CR & TURNER C (1968) Dimensions of organization structure. *Admin. Sci. Q.* **13**, 65–105.
20. PUGH DS, HICKSON DJ, HININGS CR & TURNER C (1969) The context of organization structures. *Admin. Sci. Q.* **14**, 91–114.
21. WELLS LT (1974) National policies in an international industry: the Europeans and the automobile, in *Big Business and Goverments in Western Europe* (Ed. VERNON R), Harvard University Press, Cambridge, MA.

ADDRESS FOR CORRESPONDENCE: *Professor R Mansfield, UWIST, Dept of Business Administration and Accountancy, Friary Building, 22, The Friary, Cardiff CF1 4JB, UK.*

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 139–144
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0139/$02.00/0

Some Initial Specifications for a Strategic Information System

KJ RADFORD

York University, Ontario

(*Received October* 1977)

Much of the discussion and published literature on information systems to date has been about systems concerned with the internal administration and operations of an organization. Considerably less attention has been given to the information requirements of the more strategic managerial activities of planning and policy making, in which the future directions and objectives of the organization are decided. Internal information systems in many organizations have now grown to maturity. At the same time, managers have become aware of the need for an additional component of the overall information system to support their more strategic activities. This paper considers the broad characteristics of a strategic component of an organizational information system and derives some initial specifications from which a more comprehensive design can be developed.

A FIRST consideration in the design of any information system is that it serves and supports the managerial activities of the organization of which it is an integral part. These managerial activities have been classified under three headings by Anthony [2], as follows:

Strategic planning. Deciding on the objectives of an organization, on changes in these objectives, on the resources used to attain these objectives and on the policies that are to govern the acquisition, use and disposition of these resources;

Management control. Assuring that resources are obtained and used effectively and efficiently in the accomplishment of the organization's objectives;

Operational control. Ensuring that specific tasks are carried out effectively and efficiently.

Most existing information systems have been designed and implemented to serve the second and third of these sets of managerial activities. Their impact has therefore been mainly on routine administrative and operational procedures within the organizations concerned. There have been some references in the literature to the need for information systems to support the more strategic activities of planning and policy making [6, 7, 8]. However, it is only recently that those concerned with the design and implementation of information systems have begun to consider the desirable characteristics of a strategic component of the organizational information system. Such a strategic component would be a natural counterpart to the more familiar internally-oriented system.

The relationship between the internal and strategic components of the information system can be illustrated by analogy to the running of a ship. The captain stands on the bridge and receives reports from time to time on the internal operations of the ship (such as the engine room and the passenger accommodation). These reports emanate from what can be regarded as an internal information system. The captain uses these reports to assess the efficiency and effectiveness of the internal operations and also as the basis for any necessary adjustments to the allocation of resources between these operations. At the same time, however, he also receives reports from the lookouts scanning the external environment of the ship. He uses these reports in setting the future directions of his ship towards the achievement of its goals and objectives.

TABLE 1. DETAILED CHARACTERISTICS OF AN ORGANIZATIONAL ENVIRONMENT

Participants	—objectives, goals and intentions
	—value system and beliefs
	—available courses of action
	—preferences for future outcomes
	—information available
Social elements	—relative power of participants to influence future outcomes
	—widely held values and standards of behaviour
	—relationships between participants
	—commitments resulting from past actions
	—laws, policies, guidelines, rules
Technological elements	—physical factors, geography, climate etc.
	—available technology
	—economic factors
	—human engineering factors
	—characteristics of organizations
Natural phenomena and events	—truly natural events, weather, scarcity of resources etc.
	—quasi-natural events (the unintended consequences of others' actions)

There are many differences between the characteristics of the strategic and internal components of an organizational information system. However, both conform to the description of an information system by Norbert Wiener [12] as a "... system for the acquisition, use, retention and transmission of information." The detailed design of each component depends on the managerial activities that it supports and on the type of information with which it must deal. Both these factors must therefore be taken into account in formulating the initial specifications for a strategic information system.

THE NATURE OF STRATEGIC INFORMATION

Most of the strategic activities of an organization involve its relationship with its external environment. This environment normally contains other organizations, individuals and entities, each of which is striving to move towards future situations that it finds preferable to its existing state. These organizations, individuals and entities are usually called 'participants' in the environment.

The environment has characteristics or elements that have been classified under three headings [9]: (a) *social elements*, concerned with relationships between all the participants in the environment, (b) *technological elements*, that may impose constraints on the range of courses of action available to the participants, and (c) *natural phenomena and events* that may affect the future state of the environment. Information on these characteristics of the environment and on the other participants is of vital importance in strategic planning. Detailed categories in which this information is required are shown in Table 1.

The environment of a modern organization is very wide-ranging. Portions of it may be relatively unfamiliar to many of those who work in the organization. The environment may contain a large number of sources of information and each source may provide material on one or more of a wide variety of subjects that are important to strategic managerial activities. It is not possible in many practical situations to monitor all these sources all of the time. Strategic information that is acquired by an organization is therefore often fragmentary, equivocal and less well-defined than that relating to its internal operations and administration. For this reason, the task of interpreting the meaning of strategic information is correspondingly greater than that of placing internal information in context.

Strategic information is mainly descriptive and non-quantitative. It often consists of opinions and assessments that are highly subjective. By comparison, the data in the internal information system is usually much more 'factual' in nature.

The environment of most modern organizations is not static. It has been described by Emery and Trist [3] as dynamic and turbulent.

Turbulence arises not only from rapid changes in the elements of the environment but also from the interaction between the participants. The existence of turbulence in the environment lays stress on the need for constant scanning of its characteristics in order that changes that are important to strategic management can be detected and brought to the notice of responsible decision makers.

Some strategic managerial activities are concerned also with information that exists within the organization itself. It is interesting to note in this respect, that an organization can be thought of as having an internal environment. At one stage lower in a hierarchical order, parts of an organization can be regarded as entities in their own right. The interactions between these semi-autonomous components of the organization can be regarded as a strategic problem that can be treated in much the same way as one arising in the external environment concerning the organization as a whole. Furthermore, actions and events in an internal problem may have a bearing on an external situation. The strategic information system must therefore be concerned with such information as arises within the organization as well as that originating outside it.

SOURCES OF STRATEGIC INFORMATION

Strategic information falls into four categories: (a) that which is freely and openly available, (b) information that is deliberately provided by a participant, (c) information that is held confidential by a participant, and (d) information on one's own organization that is held confidential. By far the greatest proportion of the strategic information needed by an organization is available from open sources. Observation of the activities of others and of trends in the environment by an informed individual is normally the most fruitful method of gathering this information. It is usually a mistake to place great reliance on clandestine methods of acquiring strategic information. The accent on secrecy involved in using such methods often impairs critical judgement in interpretation of the material that is collected [13].

A study of managers working in organizations showed that they rely heavily on personal contacts and networks of communication for gathering strategic information [1]. Customers and those having no regular relationship with the organization were found to be the two most important sources of information about the external environment. The study showed also that only about half of the strategic information used by managers was directly sought by them. The remainder was presented unsolicited by persons who knew of the recipient's interests and had some incentive to satisfy his needs. These findings seem to indicate that managers are not used to obtaining strategic information from a formally organized system. It may be that this is due to the fact that few such systems exist. In any case, the reluctance or willingness of managers to reveal and share information that they consider sensitive are important factors in determining the success of the strategic component of an organizational information system.

ESSENTIAL FUNCTIONS IN A STRATEGIC INFORMATION SYSTEM

The essential functions of the strategic component of an information system are very close to those described originally by Wiener [12]. They can be discussed conveniently under the headings: (a) acquisition, (b) interpretation, (c) storage and retrieval, and (d) presentation [10].

The acquisition function

Individuals and organizations are continually bombarded with strategic information. The number of sources of strategic information, the volume of information and the variety of subjects to which the information refers may be very great. Whereas it may be most desirable that strategic information be gathered on all subjects and in all areas that have a bearing on strategic decisions in an organization, this is usually impossible in practice due to constraints on the time and resources available for this activity.

Acquisition of strategic information is therefore usually carried out in a series of modes that covers a broad spectrum of activities. Three of these modes that illustrate the range of these activities are:

General scanning, which is a wide ranging and non-specific search procedure, usually

according to some predetermined pattern. It covers a wide variety of subjects and sources. It is, generally speaking, the equivalent of the activity of the scanning radar on a ship.

Directed scanning, in which scanning activity is concentrated on a particular area of the environment in which it appears some process of change is under way. It is the rough equivalent of the use of a zoom lens in photography.

Active search, which consists of a deliberate effort to seek out information on a specific subject or issue.

These three modes of acquisition of strategic information are complementary and each has its role according to the situation facing the organization at the time. General scanning is the most economic method of covering a wide range of sources and subjects, but the information it produces may not be sufficiently specific to a particular problem. Directed scanning is more concentrated, but the number of subjects that can be treated in this way is usually limited. Furthermore, the diversion of resources away from general scanning may result in areas that later turn out to have been important being left uninvestigated over longer periods of time. Active search is often the most effective method of gathering strategic information in specific areas, but it is also by far the most costly.

Management of the acquisition function is one of the most important aspects of the strategic information system. It consists of a continuous process of choosing a mix of acquisition modes that provide an adequate flow of strategic information to the responsible decision makers involved in planning and policy making activities. The process of choice of a mix of acquisition modes is aided by the establishment of what have been called 'criterion checks' [4]. The use of criterion checks involves the continuous monitoring of indicators and standards, so that any deviation that is detected can provide early warning of a possible change in conditions in the area surveyed. The deviation in normal values or patterns is then taken as a signal that a more concentrated acquisition mode should be used in the area of possible change.

Changes in acquisition mode can also be brought about by incidents other than the detection of change in the general scanning activity. For example, individual managers involved in a specific decision problem often have needs for immediate and concentrated active search as a result of their analysis of that problem. The input of such managers into the process of choice of acquisition modes often has a significant bearing on the effectiveness of the process of gathering strategic information.

The interpretation function

The purpose of the interpretation function in a strategic information system is to assess the significance of information that has been acquired. This significance may be apparent from changes and trends that are detected by comparison of the new information with that which has been gathered previously. It may, on the other hand, result from the recognition that a particular piece of information completes a pattern that had not hitherto been noticeable.

Interpretation of strategic information is inseparable from the process of acquisition. It cannot be done effectively out of context or in a partial vacuum. Managers involved in decision processes are often the best interpreters of newly-acquired information. For this reason, the basic information that has been collected in the acquisition function should always be available to those involved in the related decision processes. The best results and economy of effort are usually obtained by close and continuing cooperation between managers and their colleagues working in the interpretation function of the strategic information system. Such cooperation also provides some safeguard against distortion and misperception of the meaning of strategic information that may arise from a number of behavioural and group-interactive phenomena [5, 11].

The storage and retrieval function

The purpose of the storage and retrieval function in the strategic information system is to provide facilities for the retention of information and its later retrieval. The facilities for storage and retrieval in a strategic information system consist of the following components; (a) a series of storage locations in which documents and information are stored in various forms, (b) a system of communication with the storage media that allows users to update and

Vol. 6
1978

assess their contents, and (c) a system of organization of the stored information that contributes to its effective and efficient retrieval.

The key to an effective and efficient storage and retrieval system is in accurate, consistent and precise descriptions of the material that it contains. The type of organization, classification and indexing system that it is desirable to use with a body of information clearly depends on how well the information is described and structured. If the information is ill-structured, the system should have the nature of a general *directory*. A directory should encourage exploration of the body of information rather than the narrowing down of the area of search. As the information content of the store becomes better structured, a *subject index* or a *catalogue* can be produced. A further step in the classification of the information is the construction of a *thesaurus*, which contains linkages and cross references between the items in the store. As the body of strategic information in an organization becomes more sophisticated so too does the manner in which it can be organized and indexed. A more sophisticated method of indexing information leads to more complex and fruitful search procedures so that, with cooperation, the activities of the user and the organizer of the strategic store build upon each other.

The presentation function

The presentation function in the strategic information system is concerned with making available to managers and others responsible for decisions the information that is pertinent to their current work. The wide variety and scope of the information in a typical strategic information system precludes presentation of its entire content to all its users. A way must be found of bringing the details of the information available to the attention of the users in a manner designed to attract their interest. The chosen method must also provide users with the information they require with the least possible diversion of effort from their own tasks and responsibilities.

One of the most important aids to presentation of the content of a store of strategic information is the availability of a series of comprehensive directories and catalogues and of a competently compiled thesaurus. The aim in this most general method of presentation of the data is to display the contents in a form that is of immediate assistance to users in their search for information. This objective can be met over a period of time only if the directories and catalogues and the thesaurus are continually kept up-to-date.

Another method of presentation of strategic information consists of the establishment of an information display room located centrally in the organization. Display rooms of this sort are common in military command centers. Their purpose is to bring strategic information (and also information from the internal system) that is relevant to current decision situations to the notice of the decision makers in as comprehensive and interesting a way as possible in a short period of time. This is achieved by the use of display techniques that integrate information from a number of fields and that allow concentration on specific issues and situations. The use of modern video display technology allows rapid and flexible movement through a number of subjects. It also facilitates continuous updating of the information displayed.

The emphasis in a display room should be on interaction between the users and the store of strategic information. The intuitive reactions of managers can then be incorporated into the interpretation of the strategic information acquired by the organization. In the course of this interaction, information that has been acquired by individuals and that has not yet been incorporated into the strategic information system may become available for more general distribution and use.

AN INITIAL SET OF SPECIFICATIONS FOR A STRATEGIC INFORMATION SYSTEM

The foregoing discussion provides a basis upon which some broad initial specifications for the strategic component of an organizational information system can be derived. These initial specifications can be summarized as follows:

1. The strategic component of the information system must be designed in a manner appropriate to the handling of information that is: (a) mainly external to the organiza-

tion, (b) often fragmentary, equivocal and ill-defined, (c) mainly descriptive and non-quantitative, and (d) comprised of highly subjective opinions and assessments;

2. The design of the strategic component of the information system must be such as to support strategic managerial activities such as high-level decision making, planning and policy making;

3. The design of the strategic component must be compatible with the pattern and structure of these strategic managerial activities and must be responsive to changes in these patterns as they occur;

4. The strategic information system must incorporate the functions of; (a) acquisition, (b) interpretation, (c) storage and retrieval, and (d) presentation;

5. The design of the system incorporating these functions must be decided cooperatively by the information system specialists and the managers involved in the strategic activities that the system supports;

6. The acquisition function must provide means for information gathering in a number of different modes and a means of choosing a mix of these modes suitable to any situation, based on the characteristics of that situation;

7. The interpretation and acquisition functions must be closely interrelated. The interpretation function must have in-built features that provide safeguards against distortion and misperception of the meaning of strategic information;

8. The storage and retrieval function must provide; (a) a series of storage locations, (b) a system of communication with the storage locations, and (c) a system of organization of the stored material that allows efficient and effective retrieval;

9. The presentation function must be designed so that strategic information is made readily available to managers requiring it in their work and be such as to encourage interaction between these managers and the system.

Considerably more work is needed to expand these initial, broad specifications into a form upon which a detailed design can be based. The need for this work is urgent, taking into account the relative lack of support of strategic managerial activities by present-day information systems. Support of these activities is essential in view of the increased degree of interaction of modern organizations with their external environment.

REFERENCES

1. AGUILAR FJ (1967) *Scanning the Business Environment.* Macmillan, New York.
2. ANTHONY RN (1965) *Planning and Control Systems: A Framework for Analysis.* Graduate School of Business Administration, Harvard University, MA.
3. EMERY FE & TRIST EL (1965) The causal texture of organizational environments. *Human Relations* **18,** 21–32.
4. FERENCE TP (1970) Organization communication systems and the decision process. *Mgmt Sci.* **17**(2), B89.
5. JANIS IL (1972) *Victims of Group Think.* Houghton Mifflin, Boston, MA.
6. KEFALAS AG (1975) Environmental management information systems (ENVMIS): a reconceptualization. *J. Bus. Res.* **3,** 253–266.
7. KING WR & CLELAND DI (1973) Decision and information systems for strategic planning. *Bus. Horizons* **16,** 29.36.
8. MASON RO & MITROFF II (1973) A program for research on management information systems. *Mgmt. Sci.* **19**(5), 483.
9. RADFORD KJ (1977) *Complex Decision Problems.* Reston, VA.
10. RADFORD KJ (1978) *Information Systems for Strategic Decisions,* Chapter 6. Reston, VA.
11. STOESSINGER JG (1974) *Why Nations Go to War.* pp. 27–30; 223–230. St Martin's Press, New York.
12. WIENER N (1969) *Cybernetics.* MIT Press, Cambridge, MA.
13. WILENSKY HL (1967) *Organizational Intelligence.* p. 66. Basic Books, New York.

ADDRESS FOR CORRESPONDENCE: *Professor KJ Radford, Dept of Administrative Studies, York University,* 4700 *Keele Street, Downsview, Ontario* M3J 2R7, *Canada.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 145–152
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0145/$02.00/0

Parametric Branch and Bound

FRED GLOVER

University of Colorado

(*Received September 1977*)

The paper describes a procedure for mixed integer programming that allows branches to be imposed 'by degrees', which can subsequently be revised or weeded out according to their relative influence. It is an adaptive approach in which the branch and bound tree can be manipulated and restructured. The approach also yields measures of the costs of imposing the branches that lead to integer solutions, thus providing a built-in form of sensitivity analysis for evaluating the effect of integer restrictions.

1. INTRODUCTION

THIS paper introduces a parametric branch and bound (B&B) procedure that has greater flexibility than ordinary B&B. Once a branch is taken in ordinary B&B, it is largely irrevocable—i.e. all descendant branches must inherit the limitations imposed by their predecessors. In parametric B&B, a descendant branch may partly or wholly undo an antecedent branch. Moreover, once a feasible mixed integer programming (MIP) solution is obtained, then the 'actual' branches that achieved this solution can be identified by weeding out uninfluential branches created during the solution process.

The strategy underlying parametric B&B is to incorporate variables and constraints into the objective function in a manner resembling 'multi-objective' or 'goal programming' approaches [1, 2, 9]. The process is especially direct for 0–1 MIP problems, which do not require the creation of additional variables. However, even in the general case, all calculations can be carried out relative to a compact basis that is the same size as for the original problem.

The use of weighted variables and constraints in the objective function also bears a connection to 'Lagrangean' approaches [3, 6, 10]. However, in contrast with Lagrangean techniques, the weights are not designed to solve a dual problem, but rather entirely overshoot dual feasibility. The process may be viewed as that of constructing tentative duals (at least implicitly), guided by considerations relevant to the B&B setting. Also, in contrast to standard B&B and its exploitation of duality, parametric branches are conveniently handled by postoptimizing with the primal simplex method, whereas ordinary branches are often preferably handled by postoptimizing with the dual simplex method. Instead of imposing a branch either fully or not at all, parametric B&B allows one to impose branches 'by degrees'. This is extremely important for enabling branching alternatives to be carefully analyzed—and revised, if desirable—at later stages of the tree.

By its nature, parametric B&B yields information about the cost of imposing branches that lead to integer solutions. This information can then be used in a sensitivity analysis for evaluating the significance of integer restrictions. This type of analysis can also be coupled with an approach that attaches penalties to deviations from constraints. Such an approach yields a combined sensitivity analysis characterizing the influence of integer assignments on problem constraints.

2. PARAMETRIC B&B FOR 0–1 PROBLEMS

The basic ideas of parametric B&B are relatively straightforward for 0–1 MIP problems,

and we first examine them in this setting. For our purposes, the 0–1 problem will be written

$$\text{Minimize } x_0 = \sum_{j\in N} c_j x_j \qquad (2\text{-}1)$$

$$\sum_{j\in N} a_{ij} x_j \le b_i \qquad i\in M \qquad (2\text{-}2)$$

$$x_j \ge 0 \qquad j\in M \qquad (2\text{-}3)$$

$$1 \ge x_j \ge 0 \qquad j\in J \qquad (2\text{-}4)$$

$$x_j \quad \text{integer} \qquad j\in J \qquad (2\text{-}5)$$

where J is the index set of the integer variables and is a subset of N. In our discussion of 0–1 problems, we will confine ourselves to elaborating the principal ideas of parametric B&B, together with illustrating the more rudimentary types of considerations. In fact, most of the implementation aspects of the 0–1 case require no commentary other than to indicate the relevant decision alternatives. Subsequently, more advanced aspects of implementation will be introduced for the general MIP case, and then linked to the 'sensitivity analysis' framework.

In the 0–1 setting, consider any two disjoint subsets J_0 and J_1 of J and the 'contrived' objective function

$$\text{Minimize } z_0 = d_0 + \sum_{j\in J_0} d_j x_j - \sum_{j\in J_1} d_j x_j \qquad (2\text{-}1')$$

where all of the d_j are positive, and where

$$d_0 = \sum_{j\in J_1} d_j.$$

By the form of z_0, if there exists a feasible solution to the LP problem (2-1)–(2-4) that satisfies $x_j = 0$ for $j\in J_0$ and $x_j = 1$ for $j\in J_1$ then this will be an optimal solution to the LP problem in which (2-1′) replaces (2-1). (In fact, this assignment of values to the x_j for $j\in J_0 \cup J_1$ is uniquely optimal when feasible, and hence must occur at an LP extreme point. Note that this gives a rather simple proof of the fact that feasible integer assignments are unique extreme points of the pure 0–1 problem and occur at one or more extreme points of the mixed 0–1 problem.)

In addition, by the definition of d_0, it follows that $z_0 \ge 0$ for all feasible LP solutions, and $z_0 = 0$ only for a solution that yields the assignment $x_j = 0$ for $j\in J_0$ and $x_j = 1$ for $j\in J_1$. Consequently, the creation of a composite objective function

$$\text{Minimize } u_0 = x_0 + z_0 \qquad (2\text{-}1'')$$

assures by the non-negativity of z_0 that

$$\text{Min } u_0 \ge \text{Min } x_0.$$

Drawing on the fact that the constraint $z_0 = 0$ is equivalent to the 'partial assignment' $x_j = 0$, $j\in J_0$ and $x_j = 1$, $j\in J_1$ we may use the objective (2-1″) as a device to compel this partial assignment to hold. Clearly, if the parameters d_j are selected large enough, then $z_0 = 0$ must result provided the corresponding partial assignment is feasible.

Handling the composite objective function in parametric B&B

The first stage of the parametric B&B approach for 0–1 problems utilizes these ideas to influence the creation of partial assignments through manipulation of the parameters d_j. However, instead of assigning these parameters preemptively large values, more moderate values are assigned and then monitored in order to determine interactive effects relevant to the B&B setting. By the familiar Lagrangean type of argument, we may observe

$$\text{Min } \{x_0 \text{ subject to } z_0 = 0\} = \{\text{Min } u_0 \text{ subject to } z_0 = 0\} \ge \text{Min } u_0$$

and hence the composite objective function (2-1″) always yields a lower bound on the ordinary B&B objective (in which $z_0 = 0$ is explicitly imposed). This type of bound information can be used for fathoming in a manner resembling that of the ordinary B&B approach. We will discuss the way to accommodate fathomed alternatives in the parametric setting after introducing the notion of a 'parametric branch'.

0–1 *Parametric branching*

The branch step for the 0–1 parametric approach is a 'tentative' operation that either becomes consolidated or revised on the basis of information subsequently generated. Quite simply, if the branch corresponds to the assignment $x_r = 0$ then the current objective u_0 is updated to become $u_0 + d_r x_r$ and if the branch corresponds to the assignment $x_r = 1$ then u_0 is updated to become $u_0 + d_r(1 - x_r)$ or equivalently $u_0 + d_r x_r'$ where x_r' is the slack variable for the inequality $x_r \le 1$. The weight d_r is selected so that the updated representation of u_0 is dual infeasible, and therefore, allows reoptimization with the primal simplex method.

To illustrate, suppose the current LP representations of u_0 and x_r are given by

$$u_0 = 12\tfrac{1}{4} + 5\tfrac{1}{2}\, x_3 + 1\tfrac{1}{4}\, x_4 + 4x_5,$$

$$x_r = \tfrac{1}{4} - 2x_3 + x_4 + \tfrac{3}{4}x_5.$$

Then for the branch $x_r = 0$ any value of d_r satisfying $d_r > 5\frac{1}{2}/2$ will cause the new objective $u_0 + d_r x_r$ to be dual infeasible. (The ratio $5\frac{1}{2}/2$ is exactly the pivot ratio for the dual simplex method that would be identified if one were to introduce the constraint $x_r \leq 0$.) Thus, for example, selecting $d_r = 3$, we obtain the new u_0 objective

$$u_0 = 13 - \tfrac{1}{2} x_3 + 4\tfrac{1}{4} x_4 + 6\tfrac{1}{4} x_5.$$

The negative coefficient for x_3 of course signals that re-optimization may now be undertaken with the primal simplex method.

Alternatively, for the branch $x_r = 1$ a value of d_r satisfying $d_r > 1\frac{1}{4}/1$ will succeed in establishing dual infeasibility for the objective $u_0 + d_r x'_r$ where $x'_r = 1 - x_r$. (Here $1\frac{1}{4}/1$ is the pivot ratio for the dual simplex method relative to the constraint $x_r \geq 1$.) Thus, for example, selecting $d_r = 2$ we obtain the new objective

$$u_0 = 13\tfrac{3}{4} + 9\tfrac{1}{2} x_3 - \tfrac{3}{4} x_4 + 2\tfrac{1}{2} x_5.$$

Values of d_r such as those selected in these examples may not be sufficiently large to insure that the branches for $x_r = 0$ and $x_r = 1$ will utlimately be enforced. However, as previously noted, the procedure does not seek preemptive values but rather seeks to analyze the consequences of more moderate values. (In this connection, it is easily established that if a dual pivot on the constraint $x_r \leq 0$ or $x_r \geq 1$ would achieve primal feasibility in one step, then any value of d_r that exceeds the dual pivot ratio will achieve exactly this same solution in one primal iteration.)

This extremely simple type of 'local' implementation step for 0–1 parametric B&B has a direct analog in a variety of standard 'parametric' approaches for ordinary linear programming. The crucial aspect in the B&B setting is the way in which this step is used—i.e. the manner in which the parametric branches are processed to yield information and branching alternatives not available to ordinary B&B. Thus, in particular, parametric branches of the form just illustrated are not regarded as ironclad impositions nor are they initially assigned a sequential ranking. Indeed, it may well occur at a subsequent step that an 'earlier' branch will be discovered to be superfluous in terms of other more influential branches. We discuss this aspect of the approach next.

Revised and augmented parametric branches

Branches that are currently uninfluential can easily be singled out by the parametric B&B procedure as follows. Suppose the updated LP representation of u_0 gives x_r or x'_r (depending on which of the two variables was previously assigned a weight) a coefficient that equals or exceeds d_r. (The implication, of course, is that x_r or x'_r is currently nonbasic.) Then, reducing this coefficient by d_r still leaves the objective dual feasible. This means that the parametric branch *can be eliminated without changing the current LP solution.*

The step of weeding out uninfluential branches (i.e. those subject to elimination) can be postponed until a feasible 0–1 solution is obtained. Then, all variables whose parameter values can be reduced to zero may be excluded from the branching category. (An exception occurs for branches that are antecedents of a compulsory branch, as described in the next section.) The remaining variables and parametric branches may further be ranked, for example, according to the magnitude of the d_r values that are required to make the current solution optimal. In this fashion, the 'actual' branches of the B&B process, and their sequence, are decided upon at a stage in which the true significance of these branches can more accurately be assessed.

Another means for evaluating the relative influence or significance of a branch occurs when a branching variable 'resists' its weight and becomes basic. The decision must then be whether to increase or decrease d_r. (The latter may be a compounded step that both decreases d_r to 0 for the current parametric branch and then assigns d_r a positive value for the alternative branch.) By means of such decisions, earlier parametric branches may either be revised or reinforced.

The technical aspects of implementing these ideas are relatively simple, and hence, we confine ourselves to the task of discussing only the more prominent strategic considerations, deferring the particulars of implementation to the general MIP case, where they are more critical.

Compulsory branches and sequences

The minimization of u_0 as noted earlier provides a lower bound on the optimum value of x_0 subject to the partial assignment associated

with $z_0 = 0$. Whenever this bound equals or exceeds the value of x_0 for the best MIP solution currently known (which automatically occurs when the current LP solution in fact provides this best MIP solution), then all further continuations of the partial assignment are identified as unproductive in the usual B&B sense. This immediately provides the option of 'back-tracking'—i.e. of deciding which parametric branch underlying the current LP solution should be considered the 'last branch'. The observation that the determination of the last branch can be deferred to yield a more flexible variant of the LIFO procedure was first made by Tuan [11]. In the current approach, this determination can be based, for example, on the current d_j values (after weeding out uninfluential branches). Thus, in this fashion, a sequential ordering is gradually imposed on the branches.

Having identified a last branch, the alternative branch is now imposed as *compulsory*. This means that the method does not allow this branch to be reversed or discarded (following the usual backtracking rules), until the current partial assignment is fathomed. In particular, as soon as a compulsory branch is identified, all other parametric branches that are currently in force (i.e. all those for which d_j is currently positive) must be considered as prior to the compulsory branch. This does not impose any particular sequence on these prior branches (until backtracking again compels one to be identified as a 'last member'), but does impose *induced bounds* on the parameters d_j. That is, these positive d_j coefficients cannot be reduced below their current values as long as the compulsory branch remains in effect.

There are, however, three important exceptions to the strict maintenance of induced bounds. The first and obvious exception occurs by deciding upon some partial sequence for the parametric branches and employing a tree search rule that jumps back over some of the 'later' branches of this sequence. Then, the induced bounds can be disregarded for the branches thus bypassed. (The same remark holds for branches that are released or reversed as a result of carrying out a backtracking step with the LIFO rule.) The second exception occurs by redeciding the status of the branch that has been designated compulsory. That is, a compulsory branch may be revised, if the reconfigured objective u_0 that results after a series of additional iterations makes this appear eminently desirable. (Such a process must, of course, be sufficiently systematized to guard against circularity.) Finally, just as compulsory branches may be imposed as standard branches rather than parametric branches (as by the use of pre-emptive weights), so may the identified antecedents of the compulsory branches be imposed in the standard fashion. This imposition, however, does not carry with it an implied sequence for the antecedents themselves. Further, greater flexibility is achieved by monitoring the values of parameters that would suffice to maintain the imposed branches, thereby retaining the option of amending the status of these branches at subsequent stages.

3. PARAMETRIC B&B FOR GENERAL MIP PROBLEMS

Many of the same strategic notions discussed in the preceding section carry over to the general MIP problem. However, there is a very significant difference between the general problem and the 0–1 problem. Specifically, in the general case, there may not exist weights that will cause all of the branching inequalities to be satisfied. Moreover, if such weights exist, they may be very hard to find, or cause some inequalities to be unnecessarily over-satisfied, thus failing to identify integer solutions. We now examine the special techniques that permit these difficulties to be overcome and allow the general approach to be implemented successfully.

Branching schemes for the general MIP problem may be viewed as the successive imposition of upper and lower bounds on the problem variables. Thus, at any particular stage, the integer variables are governed by restrictions of the form

$$U_j \geq x_j \geq L_j \qquad j \in J \tag{3-1}$$

where U_j and L_j may represent original bounds on x_j or those currently inherited by branching. The general parametric approach seeks to incorporate such bound restrictions into the objective function in a manner that will enable the same types of evaluations and manipulations that are possible in the 0–1 setting.

Vol. 6
1978

This is accomplished by introducing non-negative variables z_j and the inequalities

$$U_j + z_j \geq x_j \geq L_j - z_j \qquad j \in J. \tag{3-2}$$

Thereupon, the z_j variables are incorporated into the minimization of x_0 to create the composite objective function

$$\text{Minimize } u_0 = \sum_{j \in N} c_j x_j + \sum_{j \in J} d_j z_j. \tag{3-3}$$

It is clear that for the weights d_j sufficiently large all of the z_j will be 0 and (3-2) will reduce to (3-1)—provided, of course, that (3-1) is consistent with the original problem constraints. Also, for $d_j > 0$ we have

$$\text{Min } u_0 \geq \text{Min } x_0$$

and

$$\text{Min } \{x_0 \text{ subject to (3-1)}\} = \text{Min } \{u_0 \text{ subject to (3-1)}\} \geq \text{Min } u_0.$$

Thus, the same bound relationships hold in the general MIP settings as in the 0–1 context upon introducing (3-2) and (3-3).

The problem at hand is how to manage the system (3-2) and (3-3) effectively. This is extremely important because otherwise, the introduction of (3-2) and (3-3) doubles the number of integer variables and adds a corresponding number of new constraints. The following approach for dealing with this problem constitutes an adaptation of procedures for 'weighted deviation' problems introduced in [7].

Let s'_j and s''_j denote slack variables for the inequalities of (3-2), yielding the equations

$$\begin{aligned} x_j - z_j + s'_j &= U_j, \\ -x_j - z_j + s''_j &= -L_j. \end{aligned} \tag{3-4}$$

Adding these two equations, we obtain

$$-2z_j + s'_j + s''_j = U_j - L_j. \tag{3-5}$$

Equation (3-5) identifies a 'primal relationship' between the variables z_j, s'_j and s''_j. On the 'dual' side, let u_j, v'_j and v''_j represent dual variables associated with the non-negativity restrictions on z_j, s'_j and s''_j respectively. (That is, the defining equations for u_j, v'_j and v''_j are respectively given by the z_j, s'_j and s''_j columns of the primal.) Then, since the coefficients of z_j, s'_j and s''_j are d_j, 0 and 0 in (3-3), it follows from (3-4) that the relationship between the columns for z_j, s'_j and s''_j can be summarized by

$$u_j + v'_j + v''_j = d_j. \tag{3-6}$$

This relationship, in conjunction with (3-5), provides the key to implementing the general paramatric B&B approach without modifying the size of the LP tableau. In particular, it is readily established that *at least one of the variables z_j, s'_j and s''_j must be basic and at least one must be nonbasic.*

Consequently, only one of the three primal variables need be explicitly included in the tableau at any given time, whereupon the form of the other variables is always known from (3-5) and (3-6). Similarly, x_j need not be included in the tableau since it is always capable of being recovered from an equation of (3-4). The fundamental relationships of the approach as they apply in the present setting can be summarized as follows. (Formal proofs of these relationships are omitted, since their derivation is a direct consequence of (3-5) and (3-6).)

(R.0)—Of the two variables (from the group z_j, s'_j and s''_j) that are not explicitly included in the LP tableau, one is currently basic and one is currently nonbasic.

(R.1)—*If the explicit variable is nonbasic*:

(a) The tableau row for the implicit basic variable is precisely the primal equation (3-5).

(b) The tableau column for the implicit nonbasic variable is the negative of the column for the explicit variable, except that the objective function coefficient for the implicit variable is d_j minus the objective function coefficient of the explicit variable, and the column coefficient that corresponds to the implicit basic variable is as given in (a).

(R.2)—*If the explicit variable is basic*:

(a) The tableau column for the implicit nonbasic variable is precisely the dual equation (3-6). Thus, the objective function coefficient is d_j and the column has unit coefficients in the positions corresponding to its two companion basic variables, with 0's elsewhere.

(b) The tableau row for the implicit basic variable is obtained by substituting the current expression for the explicit basic variable in (3-5) and giving the implicit nonbasic variable a unit coefficient [as in (a)].

We can now give the rules for implementing the parametric B&B approach that result from these relationships. After stating these rules, we will illustrate their use by numerical example.

1. The initial step solves the LP problem without the additional variables and constraints of (3-4).

2. The parametric branching step introduces the restrictions (3-4) in a tableau for which x_j is basic and does not satisfy $U_j \geq x_j \geq L_j$.

   (a) Identify the equation of (3-4) which implies $z_j > 0$ for s'_j and s''_j nonnegative (i.e. the equation corresponding to the 'violated bound');
   (b) Omit the slack variable s'_j or s''_j from the equation identified in (a), and use this reduced equation to replace x_j with z_j by direct substitution. The omitted slack variable is the implicit nonbasic variable, and slack variable of the other equation of (3-4) is the implicit basic variable. (These implicit variables are irrelevant at this point). The variable z_j is now basic.
3. The next component step of parametric branching is to create (or increase) the weight d_j for a selected variable z_j that is currently basic. To do this, add d'_j times the current z_j equation to the objective function equation, where d'_j denotes the increment in d_j, selected large enough to require re-optimization with the primal method. (The effect of this step on any currently implicit nonbasic variables is immediately given in relation (R.1)(b), since none of the implicit nonbasic variables governed by (R.2)(a) are affected except for the companion of z_j whose coefficient is always d_j.)
4. An arbitrary iteration of the primal method can be implemented by the following observations:

   (a) An implicit nonbasic variable provides an eligible pivot column if and only if its explicit companion is nonbasic and has a coefficient exceeding d_j (as a result of relations (R.2)(a) and (R.1)(b)).
   (b) If an implicit nonbasic variable is selected for the pivot column, then it can replace the previous explicit variable, which now becomes implicit.
   (c) No implicit basic variable need be considered in determining the pivot row unless both slack variables s'_j and s''_j are basic. If an implicit basic variable provides the pivot row, it replaces the currently explicit basic variable as the new explicit variable.

Condition 4(c) is a direct consequence of (3-5) and implies that once a slack variable of (3-4) first becomes implicitly basic, it remains irrelevant unless the other slack becomes basic—or until a new branch is introduced by tightening one of the U_j or L_j values. Branching can occur only if both of the slacks s'_j and s''_j are basic and positive (since otherwise x_j would be assigned an integer value or lie outside one of its provisional current bounds) and thus, in this case, one can simply jettison all three of the s'_j, s''_j and z_j variables [after recovering x_j from (3-4)] and start again with Step 2.

In accordance with these observations, if either U_j or L_j is a 'true' bound for x_j, then this bound can be rigidly enforced simply by rendering z_j an ineligible pivot column (e.g. by increasing d_j) whenever the slack variable associated with the true bound is nonbasic. (If z_j becomes basic when the alternative slack is nonbasic, (3-5) implies that x_j must satisfy its appropriate bound, and condition 4(c) assures that the critical slack will not become nonbasic unless z_j is nonbasic.)

We now illustrate these considerations with the following numerical example.

Example

Consider the MIP problem represented by the following 'condensed' tableau format, where x_1 and x_2 are the integer variables.

		x_1	x_2	x_3
$x_0 =$	0	−9	−4	7
$x_4 =$	7	−5	(−2)	3
$x_5 =$	12	1	−3	−3

The problem constraint equations are read directly from the tableau (e.g. $x_4 = 7 - 5x_1 - 2x_2 + 3x_3$). Pivoting on the circled element with the primal simplex method yields the optimal LP tableau

		x_1	x_4	x_3
$x_0 =$	−14	1	2	1
$x_2 =$	$3\frac{1}{2}$	$-\frac{5}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$
$x_5 =$	$1\frac{1}{2}$	$\frac{17}{2}$	$\frac{3}{2}$	$-\frac{15}{2}$

Suppose now that we wish to branch parametrically on the inequality $x_2 \leq 3$ by Step 2 of the preceding rules, thus obtaining $U_2 = 3$ and $L_2 = 0$ (where $L_2 = 0$ represents a true lower bound). The first equation of (3-4) implies $z_2 > 0$ and yields $z_2 = x_x - U_2$ upon omitting

Vol. 6
1978

the slack variable s_2'. Thus, the tableau equation for z_2 is given by

		x_1	x_4	x_3
$z_2 = x_2 - 3 =$	$\frac{1}{2}$	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$

Next, by Step 3, we add a sufficient multiple of this equation to the objective equation to create dual infeasibility. Any value exceeding the 'dual pivot ratio' 1/(5/2) will do, and we select a multiple of 1 to give $u_0 = x_0 + 1z_2$ or

		x_1	x_4	x_3
$u_0 =$	$-13\frac{1}{2}$	$-\frac{3}{2}$	$\frac{3}{2}$	$\frac{5}{2}$

Thus, putting Steps 2 and 3 together (replacing the x_2 equation by the z_2 equation and replacing x_0 by u_0) we obtain the tableau

		x_1	x_4	x_3
$u_0 =$	$-13\frac{1}{2}$	$-\frac{3}{2}$	$\frac{3}{2}$	$\frac{5}{2}$
$z_2 =$	$\frac{1}{2}$	$\textcircled{$-\frac{3}{2}$}$	$-\frac{1}{2}$	$\frac{3}{2}$
$x_5 =$	$1\frac{1}{2}$	$\frac{17}{2}$	$\frac{3}{2}$	$-\frac{15}{2}$

The primal simplex method now pivots on the circled element to obtain

		z_2	x_4	x_3
$u_0 =$	$-13\frac{4}{3}$	$\frac{3}{3}$	$\frac{9}{3}$	$\frac{8}{3}$
$x_1 =$	$\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{3}$	$\frac{3}{3}$
$x_5 =$	$\frac{16}{3}$	$-\frac{17}{3}$	$-\frac{1}{3}$	$-\frac{12}{3}$

Since optimality is re-established in a single pivot, this is the same tableau that would have been obtained by a 'dual' pivot on the constraint $x_2 \leq 3$ (except for the z_2 column). With z_2 and s_2' both nonbasic (the first explicitly, the second implicitly), we may infer that the current value of x_2 is 3 (as seen by (3-4) with $U_2 = 3$). Consequently, at the present stage the parametric branch has achieved the same effect as an ordinary branch, although we do not assign the branch a sequential 'slot' and leave it open to subsequent revision.

Suppose that $x_1 \geq 1$ is selected to provide the next parametric branch. By Step 2, the relevant expression for z_1 is $z_1 = L_1 - x_1 = 1 - x_1$ giving the tableau representation

		z_2	x_4	x_3
$z_1 =$	$\frac{4}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$-\frac{3}{3}$

By Step 3, we again seek to add a sufficient multiple of z_1 to the current objective function to produce a dual infeasibility. Since the implicit nonbasic variable s_2' has a coefficient which is the negative of the explicit z_2 coefficient (i.e. $-\frac{2}{3}$), we must take the effect of s_2' into account in the manner noted parenthetically in Step 3. By relation (R.1)(b), the current objective function coefficient of z_2 is $1 - \frac{3}{3} = \frac{2}{3}$. This yields the 'expanded' current u_0 and z_1 rows as follows

		z_2	x_4	x_3	s_2'
$u_0 =$	$-13\frac{4}{3}$	$\frac{3}{3}$	$\frac{9}{3}$	$\frac{8}{3}$	$\frac{2}{3}$
$z_1 =$	$\frac{4}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$-\frac{3}{3}$	$-\frac{2}{3}$

From this, we see that a multiple of 2 suffices to produce dual infeasibility thereby giving the new objective function representation

		z_2	x_4	x_3	s_2'
$u_0 =$	$-12\frac{1}{3}$	$\frac{7}{3}$	$\frac{11}{3}$	$\frac{2}{3}$	$-\frac{2}{3}$

Note that the 'expansion' used in this example was not really necessary since the fact that z_2 has a positive coefficient immediately implies that we consider the implicit variable instead of—rather than in addition to—the explicit variable.

Because s_2' gives the eligible pivot column, we replace z_2 by s_2' (so that z_2 now becomes implicit) to yield the tableau

		s_2'	x_4	x_3
$u_0 =$	$-12\frac{1}{3}$	$-\frac{2}{3}$	$\frac{11}{3}$	$\frac{2}{3}$
$z_1 =$	$\frac{4}{3}$	$-\frac{2}{3}$	$\frac{1}{3}$	$-\frac{3}{3}$
$x_5 =$	$\frac{16}{3}$	$\frac{17}{3}$	$-\frac{1}{3}$	$-\frac{12}{3}$

Re-optimization with the primal method gives

		z_1	x_4	x_3
$u_0 =$	-13	1	2	1
$s_2' =$	2	$-\frac{2}{3}$	$\frac{1}{3}$	$-\frac{3}{3}$
$x_5 =$	10	$-\frac{17}{3}$	$\frac{2}{3}$	$-\frac{15}{3}$

The variables of (3-4) associated with x_1 and x_2 have integer values in this tableau, and hence so do x_1 and x_2. Specifically, direct application of (3-4) yields $x_1 = 1$ and $x_2 = 1$ (noting that s_1'' and z_2 are implicitly nonbasic). Also, since z_1 and z_2 are both in the current

basic solution, it follows that u_0 and x_0 have the same value.

We now analyze the parametric branches that provided this current integer solution. By relation (R.2)(a), the objective function coefficient of z_2 is d_2. Thus, reducing the coefficient of z_2 by d_2 (which effectively removes z_2 as a weighted variable in the objective function) still yields the same LP optimum. By our earlier observations for the 0–1 case, this means that the branching restriction that gave rise to z_2 is currently 'uninfluential', and thus we can shrink the B&B tree to the single restriction $x_1 \geq 1$.

At this stage, since the current solution is locally optimal, it is unproductive to proceed with the present line of branching. Thus $x_1 \geq 1$ is replaced by the compulsory branch $x_1 \leq 0$, and the solution process continues.

4. PARAMETRIC B&B AND MIP SENSITIVITY ANALYSIS

Two types of sensitivity analysis are available with parametric B&B as illustrated in the preceding section. The first involves identifying the 'cost' of branches that lead to particular integer solutions. For example, in the earlier numerical example, it was noted that essentially no cost attached to branching on x_2 to obtain the first integer solution. (Consequently, the integer restriction for x_2 was judged 'conditionally superfluous' and the branch was discarded in this instance.) Further, the tableau that yields the integer solution assigns z_1 an objective function coefficient of 1, which means that d_1 can be reduced from its present value of 2 to $2 - 1 = 1$ without altering the LP optimum. This places the cost of the x_1 branch underlying this solution at 1, which provides a measure of the relative significance of this branch and the integer restriction on x_1 in the current solution. Specifically, we can conclude that relaxing this branch will result in at least 1 unit of improvement in the objective function for each unit of change in the value of x_1. Corresponding interpretations apply to situations in which multiple branches underlie an integer solution.

The second type of sensitivity analysis involves the explicit assignment of penalties to deviating from particular problem constraints. Coupling this penalty procedure with parametric B&B, the interactive effect of constraint deviations and parametric branches becomes susceptible to evaluation. For example, the updated cost of a variable representing a constraint violation identifies the net effective cost of imposing the constraint relative to a given integer solution. Thus, if the constraint is in fact violated at this solution, the d_j values for the parametric branches not only reflect branching costs relative to the 'original' objective function but also relative to the penalty incurred from the violated constraint. This enables a more realistic form of problem solving in situations where constraint deviations may be tolerable at a cost. The entire analysis and solution process can be conveniently carried out by merging the compact basis procedures illustrated for the MIP problem with the more elaborate procedures of [7] to maintain the working tableau the same size as the original.

Vol. 6
1978

REFERENCES

1. Charnes A & Cooper W (1961) *Management Models and Industrial Applications of Linear Programming*, vol. 1, John Wiley, New York.
2. Charnes A, Cooper W & Ijiri U (1963) Breakeven budgeting and programming to goals. *J. Acc. Res.* **1**(1).
3. Fisher ML & Shapiro JF (1973) Constructive duality for discrete optimization. CMBSE Report 7321, Graduate School of Business, University of Chicago.
4. Gass SI & Saaty TL (1955) The computational algorithm for the parametric objective function. *Nav. Res. Logist. Q.* **2**, 39–45.
5. Geoffrion AM (1966/67) Strictly concave parametric programming, Parts I and II. *Mgmt Sci.* **13**(3), 244–253 and **13**(5), 359–370.
6. Geoffrion AM (1974) Lagrangean relaxation for integer programming. *Mathl Progmg St.* **2**, 82–114.
7. Glover F (1976) Linear programming for weighted deviation problems using compact basis techniques. MSRS 76-14, University of Colorado.
8. Glover F & Tangedahl L. (1976) Dynamic strategies for branch and bound. *Omega* **4**(5), 571–576.
9. Johnsen E (1968) Studies in multiobjective decision models. *Suden Litteratur.*
10. Shapiro JF (1971) Generalized lagrange multipliers in integer programming. *Ops. Res.* **19**(1), 68–76.
11. Tuan NP (1971) A flexible tree search method for integer programming problems. *Ops. Res.* **19**, 115–119.

ADDRESS FOR CORRESPONDENCE: *Professor Fred Glover, College of Business Administration, University of Colorado, Boulder, CO* 80309, *USA.*

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 153–160
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0153/$02.00/0

An Empirical Examination of the Dimensions of Quality of Working Life

JAMES C TAYLOR

Center for Quality of Working Life, Los Angeles

(*Received May* 1977; *in revised form October* 1977)

A factor analysis was undertaken in an effort to investigate the underlying structure of the quality of working life (QWL) construct. The responses of 95 managers to 42 items were used to derive Varimax factors obtained from rotating the first five principal components. Some affinities with existing *a priori* lists of QWL criteria or categories are noted. But the total structure derived is superior to any of these, since it simultaneously deals with separate criteria which focus on individual and collective QWL concepts. Scales are produced which are explicable on the apparent meaning of the items which cluster. Adequate orthogonality and internal consistency are reported for the scales, and these in turn lead to increased confidence in the meaning of the structure obtained.

INTRODUCTION

QUALITY of working life (QWL for short) is an idea 'whose time has come.' It has been widely accepted as a unique and useful concept by policy planners, managers, and union leaders, as well as by social scientists. The task of defining quality of working life has begun with a determination of the 'qualities' of working life—utilizing comprehensive lists of attributes presumed to conform to the experiences of people at work. Such lists of qualities attempt to define quality of working life from its input, but researchers have not ignored the question of evaluation of output (i.e. the specific value of the experience on any of those qualities). Furthermore, these listings have not yet been seen as a method for prioritizing or weighting the separate qualities, independent of conventional empirical methods of testing concept values. Instead, these steps—evaluation of specific qualities in specific circumstances—have had to await completion of the definitional step of determining the qualities themselves. So far, investigators interested in the quality of working life have responded with the *a priori* creation of lists of qualities.

Quality of working life officially passed from the initial phase of introduction and acceptance to the definitional phase, with the recent publications by Davis and Cherns [3, 4] and by Biderman and Drury [1]. These books emphasize the development of indicators for quality of working life and in so doing present the definitional work to date, as well as a forum to compare the various definitions and component lists in the field. This phase also broaches the central issue of measurement, and a review of the various lists makes it clear that only initial mapping has begun. In an editorial comment on the available lists and their contribution, Eilon has pointed out that, "One cannot discuss intelligently ways in which the quality of working life can be improved or how the possible undesirable effects of certain characteristics of technology and organizational design can be avoided without an explicit measure of the quality of working life and without being able to state whether and in what manner it has changed from one period to another" [5].

Indeed, these lists of "qualities" of working life do not provide explicit measures, but at best represent constructs from which measures

may be devised. The lists of component constructs are wide-ranging and varied. Robinson [10] compares several lists and concludes that more comprehensive lists have a wider application of some subset of their components than do the less complex and shorter lists. The components of lists vary, but nearly all make reference to wages, hours, and working conditions, power in decision making, individual development, security, social integration [8, 10]; others refer to socially relevant issues, aspects of the work, and due process at the work place as well [13].

The boundaries of QWL are treated in a subtle, often unconscious way in the recent literature. Biderman and Drury [2] discuss the concept of values as a criterion measure of quality of employment, and discuss the dilemma arising from members' commitment to the employing organization versus the time and energy available for their activities outside of work. Clearly, activities off the job are not a part of working life, but the impact of the job on social relations, and home and family life can be expected to influence employees' feelings *on* the job, as well as *about* the job. Seashore [11] offers an additional perspective of QWL by looking beyond the viewpoint of the individual organization member. He proposes that QWL boundaries be extended to enclose what society and employers consider important, as well as what the given individual believes to be important. Such a comprehensive definition blurs the distinction between quality of life and quality of working life. It takes into account both society's belief that QWL means improved working conditions and employment rates, and the employers' contention that it is primarily related to worker loyalty and productivity.

This issue of considering QWL as either an 'individual' or a 'collective' concept is not unique in the development of social indicators. Gerson [6] discusses this boundary problem of the individual versus the 'transcendental' perspective in the context of quality of life concepts. He begs the question of choosing one or the other, declaring that both are useful and should be considered complementary.

Proposed indicators of quality of working life are either the result of the various authors' personal observations and *a priori* assumptions about work, or confirmatory reviews of the literature [12]. Virtually no empirical tests of any of the various lists *per se* have been reported.

Although speculative and *a priori*, the work already done provides a basis for empirical test. The present study was conducted to examine the latent structure of the comprehensive lists available. The responses of managers to a questionnaire listing of 42 separate elements of quality of working life, were factors analyzed and examined in order to study the extent and nature of the structure underlying the list.

METHOD

Measuring instruments, such as this present measure of QWL have as their basic measuring device a series of items consisting of a question or statement coupled with some response scale. Each theoretical construct included in such an instrument can be measured by a single item or a multiple item set. Of the two alternatives, the multiple-item set is much to be preferred because of the more accurate convergence of the multiple measures on a construct when compared with any single measure. In addition the effect of random response errors in a single item measure is reduced many fold through the use of multiple measures of the same construct. Of the two strategies available for developing such multiple item sets, the theory and methods using consistency and internal correlation among the items (e.g. factor analysis) are more suitable for the problem at hand than are the approaches based on the regression analysis of items on criterion variables.

Sample

A paper and pencil questionnaire was mailed to a total of 240 technical managers. One half (120) of this number was randomly drawn from the senior production engineers and engineering managers identified from the California subscription list of a widely circulated manufacturing journal (*Factory*). The other half (120) of the sample was randomly drawn from the systems analysts and EDP (electronic data processing) managers identified on the California subscription list of a popular computer and information systems magazine (*Datamation*). These two groups of technical managers were employed by large companies (over $10 million

annual turnover) in about two-thirds of the cases, and represented California-based organizations in slightly more than half of the cases. Their organizations included financial, commercial, and manufacturing firms, as well as public and not-for-profit agencies. These managers were asked to participate in a survey study of the design aspects of work systems, being conducted by the Center for Quality of Working Life of the University of California Los Angeles (UCLA). The survey covered several other more technical facets of work system design in addition to the aspect reported here. The respondents were not asked to identify themselves or their organization on the questionnaire, although they were provided a separate response card to send in if they wished to receive a report on the results of the survey. Seventy-four respondents did return the response cards with their name and affiliation. In addition the responses of the engineering group were distinguishable from those of the EDP group through the use of different coloured questionnaires.

Ninety-five completed questionnaires were returned (a useable response rate of 40%, which is quite good by current standards of response to postal questionnaires), with only a slightly greater proportion of EDP managers responding (55% > 45%). Because the questionnaires were anonymous it is difficult to be certain about the representative character of the questionnaires returned. However, the separate response cards returned did include useful identification; and a comparison was made between that information and the original sample. A close correspondence was found between the original sample of 240 and the 74 response cards returned in terms of company size (about two-thirds in each group exceeded $10 million in annual turnover), and whether the company was California based (about half in each case). This comparison lends confidence to the assumption that the responses returned do represent the sample as drawn, and that from this sample we can generalize to other modern technical managers employed in commerce, industry, and the not-for-profit sector.

Measures

A list of 42 items was generated from a combination of conceptual categories which represent the 'individual' QWL perspective [8, 13], together with constructs from the 'collective' quality of working life perspective of employers and society at large [11].

Walton's list [13], it was felt, gave a basic framework to the constructs or categories proposed in other individual lists [8, 10], as well as including the unique construct of 'constitutionalism,' which emphasises equity and due process. Twenty-four items (measuring the categories of quality of working life exemplified by Walton) were generated. Since none of the lists recently reported included measures exclusively collective from either the employer's or the larger society's point of view, other items (following Seashore [11]) were created to measure those aspects, and they were included in the questionnaire list of 42 items.

The technical managers in the sample were asked to respond to the list of 42 items in the following manner:

"*To what extent do you think your company or organization should be concerned with the following considerations?*"

Under this general question was the list of 42 items, dealing with various aspects of organizational life (which, as described above had been chosen to represent as comprehensively as possible the items from a number of QWL lists). Each of the 42 items was paired with the five-point scale:

(1) "To a very little extent"
(2) "To a little extent"
(3) "To some extent"
(4) "To a great extent"
(5) "To a very great extent"

There was *no* implication in the question, as asked, either that the respondent's existant organizational life was deficient in these areas, or that the respondent desired some change in them. The general question was designed to assess respondents attribution of relative rating of the 42 aspects for the organization in which those respondents were employed. Those ratings were studied to determine if there was some underlying pattern or structure to be found among those aspects for the sample of respondents as a whole. That structure was in turn compared with the *a priori* frameworks previously advanced in the literature.

Table 1 displays the 42 derived items and their associated categories.

TABLE 1. FORTY-TWO ITEMS AND THEIR CONCEPTUAL CATEGORIES

Questionnaire items		'Individual' QWL categories (after Walton, [13])
(1)[1]	Good pay	1. Adequate and fair compensation
(11)	Good fringe benefits plans	
(36)	Availability to individuals of different types of financial compensation	
(2)	Convenient working hours	2. Safe and healthy working conditions
(4)	Safe working conditions	
(7)	Low work-related stress	
(12)	Low work-related disease	
(20)	Job security	4. Opportunities for growth and security[2]
(28)	Opportunities to use knowledge and skills	
(38)	Opportunities for emplyees to grow and learn	
(39)	Availability of choices between challenging jobs	
(40)	Providing opportunities for careers	
(10)	Joint management–employee consultation[3]	5. Social integration in the work organization
(15)	Minimizing racial discrimination	
(19)	Good social relations at work	
(29)	Elimination of status differences	
(31)	Joint consultation between management and employees on job content[3]	
(32)	High employee self-esteem	
(33)	Better trust among members of the organization at all levels	
(35)	Honest communication among all levels of the organization	
(41)	Minimizing sex discrimination	
(6)	Standards of fairness and equity applied to all members of the organization	6. Constitutionalism in the work organization
(21)	Promoting free speech on organizational issues	
(42)	Ensuring due process to all members of the organization	
(3)	Adequate transportation to and from work	7. Work and the total life space
(23)	Opportunity to enjoy life outside the job	
(27)	High employee interest in community affairs	
(30)	Opportunity to relate job activities to product user benefits	8. Social relevance of work life
(37)	Confirmation that work performed is socially desirable	

Questionnaire items		'Collective' QWL Perspectives (after Seashore, [11])
(8)	Minimizing time lost during working hours	Employers quality of working life
(13)	High product/service quality	
(14)	Reducing theft or sabatoge	
(16)	Loyalty to the organization	
(17)	Improving employee motivation	
(18)	Eliminating scrap and waste	
(24)	Improving productivity	
(25)	High employee morale	
(26)	Minimal abuse of equipment	
(5)	Providing available recreational facilities	Societal quality of working life
(9)	Low community unemployment rates	
(22)	Loyalty to the union	
(34)	Promoting effective and active unions	

[1] Items are numbered in accordance with their presentation sequence in the questionnaire.

[2] Items were not generated to measure Walton's third category—"Immediate opportunity to use and develop human capacities" (e.g. challenge, variety of present job)—because that construct is essentially job-related, and therefore more relevant to nonmanagement perspective, and internal distinctions within the companies in the sample rather than distinction among these companies.

[3] 'Joint consultation' is not specifically mentioned by Walton although it is implied in the 'Democracy' category of Herrick & Maccoby [8].

Vol. 6
1978

Analysis

The factor analysis reported here employed a principal components analysis and orthogonal rotation (Varimax Solution) which was not undertaken until separate analyses using the two sample components (engineering managers, $n = 42$; and EDP managers, $n = 53$) had been separately analyzed and examined, using oblique rotation. Apart from a diminution in the strength of the fifth factor these individual oblique solutions for the two subsamples varied only in detail from the combined orthogonal solution reported here. This is evidence that factor structures derived from the two samples differed little, and further, that those factors are not highly interrelated. The separate oblique solutions (the actual results are not reported here in detail) showed no consistency between the subsamples for factors above five.

RESULTS

Factor analysis

The principal components solution for the combined sample yielded an unrotated general factor accounting for 23% of the total scale variance, while the next four factors together accounted for nearly 25% additional variance. Initial communality estimates were obtained using the iterative method. Table 2 presents the results of the orthogonal rotation of those first five factors (together with the Eigenvalues and variance accounted for by the principal components before rotation). Item communalities shown in Table 2 are the results achieved after factor rotation. Factor loading of 0.385 or greater was used as the criterion for including specific items in Table 2. This loading represents the significance threshold for $P < 0.001$ with $n = 95$. Thirty-two of the 42 items achieved loadings this high or higher on the five factors. A comparison between Table 2 and Table 1 reveals some (but by no means total) correspondence between the clustering of the combined Walton/Seashore lists and the empirical results. The solution obtained shows little diffusion, with only three of the 32 items loading on more than one factor.

Factors derived

The first factor contains items which focus on what could be described as 'current issues.' The items include 'due process' and 'equity' from Walton's constitutionalism category; 'safety' and 'disease' from his working conditions category; and the current social issues of racial and sex discrimination, originally assigned by Walton to his 'social integration' category. The attention currently given to these forms of racial discrimination can be traced to the influence of Affirmative Action legislation. The due process, equity, and sex discrimination items may reflect the current prominence of the Equal Rights Amendment and related issues. The safety and disease concerns are probably the result of recent activity in OSHA (Occupational Safety and Health Act) legislation.

Items concerned with the reduction of 'theft' and 'sabotage' and 'employment abuse' were expected to fall within the "employer's quality of working life" set, yet they too loaded highly on the first factor. Worker alienation is, and has been an important issue for managers, with particular reference to negative motivation. Their concern is reflected in the "Worker Alienation Bills" introduced during recent US Congresses, and in the establishment of the new US National Center for Productivity and Quality of Working Life.

The second factor, here called 'social work environment' includes those items from the Walton social integration category—'trust', 'honest communications' and 'self esteem'—as well as two additional items on joint consultation (which belong in the broadest sense to the social integration category, but are not specifically mentioned by Walton), and the item 'free speech' which Walton assigns to his 'constitutionalism' category. Factor 3 is quite a good representation of Walton's category dealing with growth and development. The four items with the highest loadings on this factor are from the five items categorized *a priori*. Factor 4 incorporates six of the eight items assigned to the 'employer's quality of working life' in Table 1 and may therefore be considered a good representation of that collective perspective. 'Eliminating scrap and waste', although having factor loadings above 0.385 with both Factor 1 and Factor 4, has only a marginal loading on the former (0.40), and a much higher loading on the latter (0.63). It is therefore appropriately considered part of Factor 4 for which it was originally intended. 'Good social relations' loads almost as highly

TABLE 2. FACTOR ANALYSIS—QUALITY OF WORKING LIFE
(Orthogonal Rotation) Technical Designers' Views of What Their Employing Organizations Should Be Concerned With

		Factors					
		1	2	3	4	5	Communality
Factor 1—'Current issues'							
(41)	Min. sex discrimination	0.72					0.57
(4)	Safe working conditions	0.65					0.47
(12)	Low work related disease	0.64					0.48
(15)	Min. racial discrimination	0.63					0.45
(26)	Min. abuse of equipment	0.60					0.58
(14)	Reduced theft & sabotage	0.59					0.46
(42)	Ensuring due process	0.52					0.47
(6)	Standards of fairness & equity	0.49					0.39
Factor 2—'Social work environment'							
(31)	Joint consultation regarding job content		0.69				0.53
(35)	Honest communication		0.62				0.46
(10)	Joint emp./mgt. consultation		0.60				0.51
(33)	Better trust		0.60				0.56
(21)	Promote free speech		0.41				0.35
(32)	High employee self-esteem		0.40				0.29
Factor 3—'Growth and Development'							
(40)	Opportunities for careers			0.72			0.61
(38)	Opportunities to grow & learn			0.67			0.50
(39)	Choices between challenging jobs			0.65			0.57
(28)	Opportunities to use knowledge & skills			0.45			0.30
(11)	Good fringe benefits			0.44			0.29
Factor 4—'Employer's QWL'							
(16)	Loyalty to organization				0.65		0.55
(18)	Eliminating scrap & waste	0.40			0.63		0.61
(24)	Improving productivity				0.60		0.43
(8)	Min. time lost during working hours				0.50		0.42
(19)	Good social relations			0.42	0.47		0.45
(25)	High employee morale				0.42		0.39
(17)	Improved employee motivation				0.42		0.43
Factor 5—'Society's QWL'							
(9)	Low community unemployment rates					0.59	0.44
(3)	Adequate transportation to work					0.56	0.34
(34)	Promoting effective unions					0.56	0.37
(37)	Confirmation of socially desirable work					0.42	0.51
(23)	Opportunity to enjoy life outside work	0.39				0.42	0.42
(22)	Loyalty to union					0.42	0.21
	Eigenvalues	9.57	3.68	2.46	2.17	1.80	
	Percent variance	23%	9%	6%	5%	4%	

N. respondents = 95
N. items (total) = 42

on Factor 3 as it does on Factor 4, and may be an equally appropriate complement of either.

Factor 5 contains three of the four items which are designed to relate directly to societal concerns: low unemployment rates, loyalty to unions, and promoting effective unions. The other three items loading on Factor 5 ('transportation to work,' 'socially desirable work' and 'enjoying life outside work') have general relevance to Walton's 'individual' categories, but are also pertinent to the 'collective' concern of 'society's QWL.' However, in addition to accounting for the lowest variance (4% in the unrotated solution), this factor among the five, is the only one not well replicated between the two subsamples in the separate oblique solutions undertaken prior to the orthogonal rotation reported here. In the earlier comparison, the EDP managers data did show this factor quite strongly, while data for the engineering managers did so to a lesser degree.

Vol. 6
1978

The three items intended to measure Walton's first construct—'adequate and fair compensation'—did not load either separately or together in a single factor. Given the apparent contemporary emphasis on the 'hot' social issues in Factor 1, compensation may be so insignificant an issue to these managers as to be literally ignored or treated in a random fashion in this context. Other respondents (perhaps nonsupervisory employees or union officials) might well generate another factor structure in this case—one which includes 'compensation' as a separate factor.

Table 3 shows the correlations among the five scales, constructed using the items with factor loadings of 0.385 and higher. The coefficients range from 0.26 to 0.53 (Median = 0.39) and can be considered adequate evidence for separate scales.

Although the total factors were subjected to orthogonal rotation, these subsets relate at a level as high as 0.53 between the first factor—'current issues'—and factor four—'employer's quality of working life.' In all, the pattern in Table 3 seems appropriate and the correlations are low enough to allow the conclusion that the factors are measuring different concepts.

Internal consistency reliability

The measure of inter-item, or internal, consistency of the five factor scales reveals the degree to which the different items within each factor give the same results. Items which give the same result are said to be measuring the same characteristic, and the combination of a set of items in each factor is justified. 'Coefficient Alpha' was applied in the present case and the results are reported in Table 4. The internal consistency of scales is measured by Alpha Coefficients, which generally increase with the number of items in the scale. Scale 5 with the lowest Alpha (0.71), and six items, must be considered a less reliable measure in view of the number of items it contains [9]. Factor 5 was the only factor not confirmed in both individual factor analyses, being a stronger factor for the EDP managers group.

Table 3. Intercorrelations among the five factor scales

Scale	1	2	Scale 3	4	5
1	—	0.40	0.31	0.53	0.26
2		—	0.49	0.49	0.45
3			—	0.38	0.37
4				—	0.32
5					—

Table 4. Internal consistency reliability for the five factor scales

Factor scales	Alpha (α)	Number of component items
1	0.85	8
2	0.82	6
3	0.78	5
4	0.81	7
5	0.71	6

Item-to-total-score analysis in one of the more frequently used methods of determining the scalability of individual items. Correlations between individual items and total factor scales constructed from Table 2 above (minus the particular item in question) were calculated for each relevant item-total pair. If a scale is homogeneous, we would expect the item-total correlations to be large and positive. Each item-total correlation coefficient would reflect the discriminability of that item or its relative strength in contributing to the overall measurement of the factor index. This item analysis, presented in Table 5, reveals acceptable item-total correlations for all items and factor scales.

DISCUSSION

Data presented in the previous section show that QWL categories aggregated from the *a priori* listings of a number of investigators, can be approximated by examining the network of

Table 5. Item to Total Correlations

Items*	Scales 1	2	3	4	5
1	0.75	0.68	0.69	0.48	0.40
2	0.58	0.55	0.63	0.63	0.52
3	0.63	0.63	0.58	0.66	0.46
4	0.60	0.66	0.43	0.53	0.48
5	0.63	0.48	0.40	0.44	0.41
6	0.59	0.46	—	0.52	0.38
7	0.57	—	—	0.52	—
8	0.52	—	—	—	—
Median	0.60	0.59	0.58	0.52	0.44

* The order of the items for each scale is the same order in which the items are arranged in Table 2.

relationships among empirically measured subscales. The factors derived from the present study do not validate the *a priori* listing of any one author. The myriad lists currently available in general require further elaboration; their definitions still beg semantic questions, and the resulting criteria will inevitably overlap [5]. This exploration demonstrates that separable factors can be derived, that the resulting index scales have satisfactory psychometric characteristics, and that these factors relate generally as predicted to a combination of the criteria proposed by the various authors.

This analysis will need to be repeated with a more systematic and comprehensive sample selection. The strength of the first four factors and their component items, however, strongly suggests that similar factors will result from wider testing.

REFERENCES

1. Biderman AD & Drury TF (1976*a*) *Measuring Work Quality for Social Reporting*. John Wiley, New York.
2. Biderman AD & Drury TF (1976*b*) Social and moral qualities of work as social indicators in *Measuring Work Quality for Social Reporting* (Eds. Biderman AD & Drury TF). John Wiley, New York.
3. Davis LE & Cherns AB (1975*a*) *The Quality of Working Life*, Vol. 1: *Problems, Prospects and State of the Art*. Free Press, New York.
4. Davis LE & Cherns AB (1975*b*) *The Quality of Working Life*, Vol. 2: *Cases and Commentary*. Free Press, New York.
5. Eilon S (1976) Editorial: the quality of working life. *Omega* **4**(4), 367–373.
6. Gerson EM (1976) On the quality of life. *Am. Sociol. Rev.* **41**, 793–806.
7. Harman HH (1976) *Modern Factor Analysis* (3rd Ed.). University of Chicago Press.
8. Herrick NQ & Maccoby M (1975) Humanizing work: a priority goal of the 1970's in *Quality of Working Life*, Vol. 1 (Eds. Davis LE & Cherns AB). Free Press, New York.
9. McKennell A (1971) Attitude measurement: the use of coefficient alpha with cluster or factor analysis. *Sociology* **5**, 227–245.
10. Robinson JP (1976) Some approaches to examining quality of employment indicators for disaggregate segments of the work force, in *Measuring Work Quality for Social Reporting* (eds. Biderman AD & Drury TF). John Wiley, New York.
11. Seashore SE (1975) Defining and measuring the quality of working life, in *The Quality of Working Life*, Vol. 1 (eds. Davis LE & Cherns AB), Free Press, New York.
12. Taylor JC (1973) Concepts and problems in studies of quality of working life. A report prepared for the Manpower Administration, US Department of Labor, Contract No. 81-06-72-09, Center for Quality of Working Life, Institute for Industrial Relations, UCLA, 67 pp.
13. Walton RE (1975) Criteria for quality of working life in *Quality of Working Life*, Vol. 1 (Eds. Davis LE & Cherns AB). Free Press, New York.

ADDRESS FOR CORRESPONDENCE: *James C Taylor Esq, Center for Quality of Working Life , Institute of Industrial Relations, University of California, Los Angeles, CA* 90024, *USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 161–171
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0161/$02.00/0

A Model for National Planning Policies

SC AGGARWAL

DJ CLARK

The Pennsylvania State University

(*Received August* 1977)

This paper attempts to develop a multiobjective model to help policy-makers choose the most appropriate mix of economic policies concurrent with the elimination of serious gaps. Each individual policy attempts to achieve a specific goal or a set of subgoals; but it may hinder the achievement of some other subgoals. To differentiate between the importance of different objectives, the analyst assigns different priorities to each of the given set of subgoals. Three sets of results are provided by three separate computer runs of this model. Each set of results corresponds to one of the three priority structures, and each run indicates quite differing underachievements or overachievements of individual subgoals. A detailed explanation for each of these deviations (gaps) from the targeted subgoals is attempted in an overall framework and some interactions between various gaps (savings gaps, foreign exchange gap, domestic growth gap, GNP gap, etc.) are also identified.

1. INTRODUCTION

NATIONAL planners require an approach for choosing the best mix of economic policies. Even large-scale econometric models, such as the Wharton Econometric Model, are not always effective, owing to rapid and wide fluctuations in the economic environment. Inflation, serious supply shortages (particularly in non-renewable materials), and the vicissitudes of the international markets and money flows have raised unusual difficulties with which planners must contend [11, p. 50].

This paper develops a model that employs the concepts of goal programming for national planning in the US. Goal programming allows complex and often conflicting subgoals to be systematized. Since multiple objectives often conflict, goal programming uses ordinal ranking of different objectives in terms of their importance. A planner can assign various combinations of rankings to a chosen set of subgoals, obtain a goal programming solution for each combination, and then select the most appealing set of rankings [7, p. 21].

The goal programming model is normative, in contrast to contemporary econometric models which are predictive in nature. This is not to suggest that these two different model-types are mutually exclusive—quite the contrary. The normative model presented here contains both exogenous and endogenous variables. A forecast of the required exogenous variables can only be provided by predictive models, and the chosen targets for the goals help quantifying the endogenous variables.

Spulber and Horowitz have described two and three gap models in a recent brilliant and comprehensive text [10]. They state that gap models help determining the actual or assumed gaps, and this way in identifying discrepancies between projections based on normative goals and the forecasts of the available resources. The two-gap model (involving savings gap and foreign exchange gap) help determining how one of these two gaps can be eliminated without widening the other. The three-gap model, in addition, includes the domestic growth gap, and attempts to eliminate the dominant gap/gaps without affecting the other gaps/gap. Further, on p. 249 [10], they state that gap models help appraising qualitative limits and determining alternative strategies about GNP,

savings, foreign capital flow, or the effective exchange rate. The goal programming model as used here in this study may be considered as an extension of the gap models, because it allows decision-makers to make changes in the national goals (as and when they decide), either in terms of the priorities of the individual goals or in the parameter values of the constraint equations. This way the goal programming model may complement the existing economic and econometric planning models.

In the following pages, while developing constraint equations of this model, we employ the concepts of Keynesian Systems of Econometric Theory in an innovative way, so as to fit these economic concepts into the framework of a goal programming model.

2. AN EMPIRICAL STUDY

The data

To design and analyze the model, empirical data available from major National Income and Products Account in the US is used. The council of economic advisors is responsible providing guidance to the Executive branch for for attaining economic stability, and a comparatively high employment level, for stabilizing foreign transactions, for regulating government expenditures, personal consumption and personal investments, for achieving planned growth in productive capacity and for keeping the personal and government savings plus exports at a predetermined level. The present model is designed to cover 1 year of the planning period. The receipt side of the National Accounts is assumed to be given. Forecasted 1974 values of these national accounts constitute the exogenous variables of the model. Given these receipts, the function of this model is to allocate the incomes among the various expenditure accounts in such a way as to obtain the highest achievement of the objectives specified by national planners. The expenditure accounts are therefore treated as endogenous variables of the model. All the estimates of revenues, personal disposable income, corporate retained earnings, GNP, capital consumption allowance and of capital stock were assumed to be reasonable and to remain fixed during the year under consideration. Therefore, these estimates can be treated as exogenous variables and are designated as b_i. The personal expenditures, government expenditures, interest payments, government subsidies, investments, change in inventories, net exports and personal savings were supposed to be governed by the priorities attached to the national goals, and in consequence these can be treated as endogenous variables and are designated as x_i. In addition to exogenous and endogenous variables, there are three coefficients that are the established parts of interlinking sectoral accounts and are designated by a_i. The list of all the variables and coefficients is given in Table 1.

SUBGOALS, CONSTRAINTS AND THE OBJECTIVE FUNCTIONS

Equilibrium level of output. For the total national economy, the basic subgoal each year is that the aggregate demand equals aggregate supply. The aggregate demand is measured by the total level of spending on goods and services in all the sectors, net of depreciation. Aggregate supply or total output is measured by gross national product less capital consumption allowances or by net national product (NNP) valued in current dollars. However, equilibrium NNP may not be equal to full employment NNP. If equilibrium NNP is less than full employment NNP, then there is deflationary spending. If the reverse is true, then there is inflationary spending [8, pp. 241–242, 181, 182, 164–165 and 357]. Both of these situations have serious repercussions on price levels and employment, and should be minimized. Attainment of this objective would then ensure a certain degree of economic stability. This relationship is expressed below where d_1^- measures deflationary spending and d_1^+ measures inflationary spending. It is assumed that inflationary spending is twice as bad as deflationary spending. Now,

(personal consumption expenditures) + (government purchases of goods and services) + (net investment expenditures) + (net exports) + $d_1^- - d_1^+$ = (full employment NNP in current dollars).

If capital consumption allowance is added to both sides, net investment becomes gross investment and full employment NNP becomes potential GNP valued in current dollars. Potential GNP is the assumed upper limit of real output of the national economy, given full employment which has been defined as 4% un-

TABLE 1. VARIABLES AND DATA

The item	Variable notation	1974 Estimate (billion $)
A. Personal expenditures and savings		
Personal expenditure on durable goods	x_1	130.2*
Personal expenditure on nondurable goods	x_2	375.2*
Personal expenditure on services	x_3	361.4*
Personal savings	x_4	74.4§
Personal transfer payments to foreigners	x_5	1.21†
Investment expenditures on residential construction	x_{21}	48.7§
Personal disposable income	b_1	900†
Interest paid by consumers (assuming that the given amount of credit in the economy does not change)	b_2	22.7†
B. Government expenditures and revenues		
Government purchases for national defense	x_6	78.75§
Government purchases for space research and technology	x_7	3.0§
Government purchases for miscellaneous functions of government	x_8	26.565§
Government purchases for international affairs and finance	x_9	0.1§
Government purchases for education	x_{10}	75.6§
Government purchases for health, education and welfare	x_{11}	56.805§
Government purchases for veteran benefits and services	x_{12}	3.465§
Government purchases for commerce, transportation and housing	x_{13}	34.86§
Government purchases for agriculture and agricultural resources	x_{14}	1.7§
Government purchases for natural resources	x_{15}	7.77§
Government transfer payments to individuals	x_{16}	83.81§
Government transfer payments to foreigners	x_{17}	25.69§
Net interest paid by government	x_{18}	14.4§
Total subsidies less surplus of government enterprises	x_{19}	10.1§
Government surplus or deficit (National Income and and Product Accounts)	x_{20}	0.00§
Personal tax and non-tax revenues	b_3	159.0†
Corporate profit tax accruals	b_4	56.2†
Indirect business tax and non-tax accruals	b_5	120.00†
Social security contributions	b_6	67.50†
C. Corporate transactions		
Investment expenditures on plant and equipments	x_{22}	153.0§
Change in business inventories	x_{23}	17.4§
Net foreign investments	x_{24}	1.0§
Export of goods and services	x_{25}	138.6*
Import of goods and services	x_{26}	138.4*
Corporate retained earnings	b_7	51.50†
Capital consumption allowance	b_8	112.00†
Capital stock at the beginning of the year (See [9], p. 25)	b_{11}	1190.2‡
D. Other miscellaneous estimates and coefficients		
Potential GNP valued in current dollars (See [4], p. 69 and [2])	b_9	1125.2‡
Potential GNP valued in constant 1958 dollars	b_{10}	884.1‖
GNP price deflator to convert 1974 dollars to constant 1958 dollars (See [4], pp. 249–250]).	a_1	0.6‡
Capital output ratio (See [6, p. 53] and [4, p. 272])	a_2	3.00‡
Desired growth rate in productive capacity (See [4], p. 27)	a_3	3.0%‡

* Source: *Survey of Current Business*, July 1974.

† This estimate has been obtained using linear regression on the actual data of the past 10 years (significance level 0.05).

‡ This estimate is from a published source, given just below the description of the variable in question.

§ This is the estimate for 1974, and it is based on the actual figure of 1973 available from *Survey of Current Business*, July 1974. The explanation, how the 1973 figure has been modified to 1974 estimate is given in the text under the subheading 'expenditure constraints..'

‖ The potential GNP from 2nd quarter of 1969, valued in constant 1958 dollars was used as a base, and then an annual growth rate of 4% was applied to this 1969 figure. The estimate of 1974 potential GNP (in constant 1958 dollars) was calculated to be equal to $884.1 billions.

employment [8, p. 181]. Any spending above potential GNP valued in current dollars is merely a function of higher prices and not the increased output. The algebraic form of this relationship is shown in equation (1) in the Appendix.

Economic costs of unemployment. Another important subgoal for a nation is the cost of unemployment. Government economists have measured the cost of unemployment in terms of a 'GNP gap.' This gap is merely the amount by which the actual GNP falls short of the potential GNP measured in constant 1958 dollars. The unemployment rate is closely correlated with the GNP gap. Minimizing the gap should therefore ensure full employment of the labor force and other productive resources. The GNP price deflator a_1 has been used to convert actual expenditures to expenditures in constant 1958 dollars, the units of potential GNP in this constraint. Here underachievement of full employment (d_2^-) is considered twice as bad as its overachievement (d_2^+):

a_1 {(personal consumption expenditures) + (government purchase of goods and services) + (net investment expenditures) + (net exports)} + $d_2^- - d_2^+ = b_{10}$. Or equation (2) in the Appendix.

Net investment for economic growth. The third subgoal for a nation with an expanding economy is its rate of economic growth which is dependent upon its net investment expenditures. Planned investment expenditures must include expenditures to replace capital stock which will be consumed during the course of the year, and also expenditures for new capital stock which will increase productive capacity commensurate with the assumed growth rate [8, pp. 164–165]. The capital–output ratio expresses the relationship between net investment and the resulting increase in the productive capacity. During the period 1966 to 1973, manufacturing capacity in the United States increased by 28% i.e. at an average annual growth rate of about 4% [6, p. 53]. For the same period, the gross domestic investment increased at an average annual rate of about 12% [4, p. 272]. Using these two rates, the capital–output ratio, a_2, works out to be nearly 3:1. If the capital–output ratio (a_2) is multiplied by the assumed growth rate in productive capacity (a_3), the result is the percentage increase in beginning capital stock (b_{11}) needed over and above depreciation replacements allowance (b_8). For 1974, a_3 has been estimated to be nearly 3% [4, p. 27], and b_{11} (at the beginning of the year) to be $1090.2 billions [9, p. 25]. This relationship is expressed below in equation (3), where d_3^- measures underachievement in growth of productive capacity, and d_3^+ measures overachievement of this very capacity. We assume that underachievement of productive capacity is twice as bad as its overachievement:

(Gross investment expenditures) + $d_3^- - d_3^+$ = (capital–output ratio) (assumed growth rate) (capital stock at beginning of period) + (capital consumption allowance). Or equation (3) in the Appendix.

Interlinking sectoral accounts (constraints). Gross National Income is allocated to three different channels. It is (1) paid to households as part of personal income, or (2) accrued by federal, state, and local governments as government receipts, or (3) retained by business as part of gross savings in the economy. The interlinking sectoral accounts trace the disposition of Gross National Income (See Fig. 1), [12, p. 66]. These accounts constitute the main model constraints and therefore are explained below both in descripitive and algebraic form.

Personal Income and Outlay Account:
(personal expenditures on durable goods) + (personal expenditures on non-durable goods) + (personal expenditures on services) + (personal savings) + (personal transfer payments to foreigners) = (personal disposable income) − (interest paid by consumers). Or equation (4) in Appendix.

Government Receipts and Expenditures Account (Federal, State and Local):
government purchases of goods and services for (national defense + space research and technology + miscellaneous functions + international affairs and finance + education + health, labor and welfare + veterans benefits and services + commerce, transportation and housing + agriculture and agricultural resources + natural resources) + (government transfer payments to persons) + (government transfer payments to foreigners) + (net interest paid by government) + (subsidies less surplus of government enterprises) + (surplus or deficit, national income and product account) = (personal tax and non-tax revenues) + (corporate profit tax accruals) + (indirect business tax and non-tax accruals) + (social security contributions). Or equation (5) in Appendix.

Savings and Investment Account:
(investment on residential construction) + (investment on plant and equipment) + (change in business inventories) + (net foreign investments) = (corporate retained earnings) + (capital consumption allowance) + (personal savings) + (government surplus or deficit). Or equation (6) in Appendix.

Foreign Transactions Account:
(exports of goods and services) = (imports of goods and services) + (government transfer payments to foreigners) + (personal transfer payments to foreigners) + (net foreign investment). Or equation (7) in Appendix.

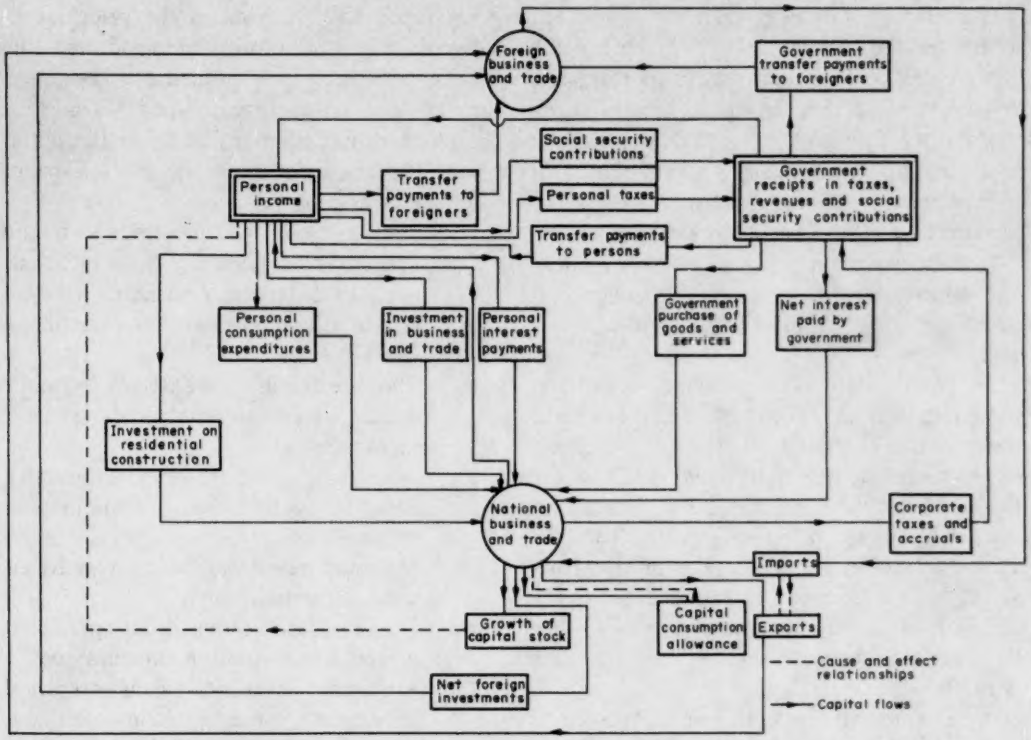


FIG. 1. *The capital flows in a nation's economy: the bases for the interlinking sectoral accounts.*

Expenditure constraints

Theoretically, there cannot be any definite limits on personal consumption expenditures and personal savings, but in practice based on the past trends economists can provide reasonable estimates of a personal consumption and savings. These estimates may be used to satisfy the requirements of a goal programming model in the form of limits (subgoals).

Corporations always prepare their annual budgets in the light of general economic conditions, and economists attempt to estimate fixed business investments and net exports for the upcoming year. This means we can again use these estimates as the limits (subgoals) in our goal programming formulation.

The limits on individual items of Federal budget are regulated from year-to-year by the joint agreement between the Congress and the President. Reaching such a mutual agreement is an extremely complex political process. Therefore, it is nearly impossible to model the process that sets limits on these government expenditures. However, once the Federal annual budget has been approved, the upper limits on each of the expenditure items is always observed, and hence these limits can be treated as subgoals for the goal programming model, as explained in the following paragraphs.

The estimates of personal consumption expenditures for 1974, (x_1, x_2 and x_3) were increased by the same percentage as the increase in GNP. Their growth was assumed to be 12% over those of 1973 [4, p. 47]. The three personal consumption expenditure constraints take the form of equations (8), (9), and (10) (in the Appendix).

The estimate of personal savings, x_4, was assumed to remain unchanged from its 1973 level; but the personal transfer payments to foreigners, x_5, were assumed to increase in accordance with the past trend determined by the linear regression. The limits on x_4 and x_5 are indicated by the constraint equations (11) and (12).

Government purchases of goods and services (for all levels of governments) were increased by 5% over those of 1973 [4, p. 45] (with the exception of x_7, x_9, and x_{14}, which were held constant). All these limits are represented by equations (13)–(22).

Government transfer payments, x_{16} and x_{17}, were increased over those of 1973 by \$14 billion [4, p. 41]. This total was proportionately divided between transfers to persons, x_{16} and transfers to foreigners, x_{17}. The net interest payments by government, x_{18} were held at 1973 amounts and the government subsidies account, x_{19} was assumed to stay put during 1974. The government surplus account, x_{20}, was reduced to zero, as a result of the increased spending. The constraint equations are (23)–(27).

The fixed business investment (new plant and equipment), x_{22}, was considered to be the major source of strength in 1974 demand, and was increased by nearly 12% over 1973 [4, pp. 43–44]. Inventory investment, x_{23}, was increased by about \$2 billion, mostly from the retail stocks of new cars. Residential construction, x_{21}, was decreased by 15% because of the tightness of money markets [4, pp. 43–44]. These assumptions are reflected in equations (28)–(30).

Since net exports (x_{25}–x_{26}) were expected to approach zero in 1974 [4, p. 46], the dollar values of exports and imports used in the model were the second quarter seasonally adjusted annual rates for 1974 [6]. Net foreign investment, x_{24}, was held unchanged from 1973, because of the devaluation of dollar. These assumptions are included in equations (31)–(33).

3. THE SOLUTIONS AND THEIR DISCUSSIONS

In a goal programming formulation the model dictates the solution in accordance with the priority structure of the subgoals setup by the national planners. In the following, three separate solutions are presented according to different priority structures of their goals. The authors have used the goal programming computer program of Lee to obtain each of the following three solutions [7, Chapter 6].

Solution 1. The planners cannot easily set definite priority structures of the national goals. To start with, they can provide one intuitively appealing set of goals, and if the model solution can show them which of the goals will be achieved, and which ones will not be achieved, the planners would definitely like to revise their initial priority structure of these goals. With each revision of the priorities of subgoals, a new solution is obtained, and this process of revising goal priorities is continued until a most satisfactory solution to most of the planners has been realized. Let the initial priority structure for run 1 be the following:

Priorities of goals (*solution* 1)

P_1: The first priority is to attain economic stability by minimizing both inflationary and deflationary spending.

P_2: Attain full employment by minimizing the 'GNP gap.'

P_3: Stabilize foreign transactions by minimizing deviations from predicted payments abroad.

P_4: Achieve assumed growth in productive capacity by minimizing deviations in either direction.

P_5: Minimize deviations from predicted investment expenditures.

P_6: Minimize deviations from predicted personal consumption expenditures.

P_7: Minimize deviations from predicted government expenditures in category-I, which includes national defense, education, health–labor–welfare, commerce–transportation–housing, and natural resources. These are assumed to be the more important areas of government involvement.

P_8: Minimize deviations from predicted government expenditures in category-II, which covers all the remaining expenditure accounts not mentioned in category-I.

P_9: Minimize deviations from predicted values of residual accounts, consisting of personal savings, government surplus, and exports.

Given these priorities, the objective function takes the form:

$$\begin{aligned}\text{min. } z = {} & P_1(d_1^- + 2d_1^+) + P_2(d_2^+ + 2d_2^-) \\ & + P_3(d_8^- + 2d_8^+ + d_{20}^- + d_{26}^- \\ & + 2d_{26}^+ + d_{28}^-) + P_4(2d_3^- + d_3^+) \\ & + P_5(d_{23}^+ + d_{24}^+ + d_{25}^+) + P_6(d_4^- \\ & + d_5^- + d_6^-) + P_7(d_9^- + d_{13}^- \\ & + d_{14}^- + d_{16}^- + d_{18}^-) + P_8(d_{10}^- \\ & + d_{11}^- + d_{12}^- + d_{15}^- \\ & + d_{17}^- + d_{19}^- + d_{20}^- + d_{21}^-) \\ & + P_9(2d_7^- + d_7^+ + d_{22}^+ + d_{27}^+) \quad (34)\end{aligned}$$

subject to all the constraints listed in the Appendix.

TABLE 2. MODELS' RESULTS FOR THE VARIABLES (billions of dollars)

Variable	Model 1	Model 2	Model 3
x_1	145.90	145.90	145.90
x_2	378.60	378.60	378.60
x_3	352.80	352.80	348.80
x_4	0.00	0.00	0.00
x_5	0.00	0.00	0.00
x_6	46.70	77.90	77.90
x_7	0.00	2.30	0.00
x_8	0.00	26.60	10.20
x_9	0.00	0.10	0.10
x_{10}	75.90	75.90	75.90
x_{11}	56.80	56.80	56.80
x_{12}	0.00	3.50	3.50
x_{13}	34.90	34.90	34.90
x_{14}	0.00	1.60	1.60
x_{15}	7.80	7.80	7.80
x_{16}	83.80	26.60	83.80
x_{17}	25.70	25.70	25.70
x_{18}	14.40	14.40	14.40
x_{19}	10.10	0.00	10.10
x_{20}	46.60	46.60	0.00
x_{21}	48.60	48.60	48.60
x_{22}	153.00	153.00	114.90
x_{23}	8.50	8.50	0.00
x_{24}	0.00	0.00	1.00
x_{25}	164.10	164.10	165.10
x_{26}	138.40	138.40	138.40

The results of the first computer run (along with those of the other two runs) are listed in Tables 2 and 3.

Solution 2. For this run, the goal priorities P_1 and P_2 of run 1 were interchanged to find out what the effects on the rest of the economy would be if full employment is assigned a higher priority, then what will be its effects on the rest of the economy. All the remaining goal priorities were kept as in Run No. 1. Since the only change in the priorities for this run is in terms of transposition of P_1 and P_2 of run 1, therefore the new list of priorities and the new objective function is not repeated here. The results of the second run also are listed in Tables 2 and 3.

Solution 3. For this run, the goal priorities 7 and 8 of run No. 1 on the objectives related to government expenditures categories-I and -II were interchanged with the priorities 5 and 6 of the objectives—investment expenditures and personal consumption expenditures respectively. The purpose of this change in priority structure was to see how personal expenditures and investment expenditures would get affected. In addition, a change was simultaneously made in capital–output ratio from 3:1 of run No. 1 to a new ratio 1:1 to explore what changes would occur if the capital–output ratio anyhow could be improved to 1:1 by way of new technology or increased productivity. This ratio change required changing the right-hand side of the constraint equation (3) to \$144.7 billion [i.e. $(1 \times 0.03 \times 1090.2) + 112 = 144.7$]. After the changes, the priorities list to solution 3 would be as follows:

Priorities of goals (*Solution* 3)

P_1: Same as of run 1.
P_2: Same as of run 1.
P_3: Same as of run 1.
P_4: Same as of run 1.
P_5: The fifth goal is that the government expenditures Category-I must take priority over investment expenditures and personal consumption expenditures.
P_6: The sixth goal becomes that the government expenditures Category-II also take priority over investment expenditures and personal consumption expenditures.
P_7: The seventh goal assumes that the estimated investment expenditures are incurred before any increase in personal consumption.
P_8: The eighth priority is assigned to personal consumption expenditures.
P_9: The same as in run 1.

As regards objective function of run 3, it need not be repeated here, because for this run the changes needed in equation (34) of solution 1 are—(i) P_5 is interchanged with P_7, and (ii) P_6 is interchanged with P_8. The results of this third run are again listed in Tables 2 and 3, like those of the other two runs.

SPECIAL COMMENTS AFTER STUDYING THE RESULTS OF INDIVIDUAL RUNS

Run 1

Net exports, $(x_{25} - x_{26})$ were determined to be \$25.7 billions, significantly above the predicted near-zero level for the year 1974. This is nothing else than the foreign exchange gap resulting from excessive oil imports. It should also be noted that a part of the government spending has been diverted from purchase of goods and services to non-GNP accounts such as transfer payments, net interest payments, and subsidies (x_{16} through x_{20}).

TABLE 3. MODELS' RESULTS IN TERMS OF GOAL ATTAINMENT (billions of dollars)

Priority	Model 1	Model 2	Model 3
P_1	Underachieved by \$210 billions—which means for attaining zero inflation the spending must be reduced by this amount.	Underachieved—but the underachivement with respect to full employment has been reduced to \$41.5 billion, compared to \$83.00 billion of the first run.	Not achieved—the inflation equals to \$210.10 billion.
P_2	Underachieved by \$83 billions, and therefore to achieve full employment this much of additional amount in constant 1958 dollars must be spent.	Not achieved—the economic stability has gone up from \$210.10 billions of 1st run to \$279.20 billions.	Not achieved—the full employment underachieved by \$83.00 billion.
P_3	Fully achieved—the limits on foreign transactions were not exceeded.	Achieved.	Fully achieved.
P_4	Fully achieved—the desired amount of \$210.10 billions of productive capacity addition was accomplished.	Achieved.	Not achieved—the productive capacity is exceeded by \$18.8 billion above its estimated limit.
P_5	Not achieved—the investment expenditures remained \$8.90 billion below their target.	Not achieved—there is no change in the underachievement of investment expenditures from \$8.90 billions of run 1.	Fully achieved.
P_6	Not achieved—the total estimated personal consumption remained \$24.30 billions below the planned target.	Not achieved—again, no change in the underachievement of personal consumption as compared to underachievement of \$24.3 billions in run 1.	Not achieved—government expenditures category-II underachieved by \$18.7 billions.
P_7	Not achieved—the category-I government expenditures stayed below their planned level by \$31.20 billions.	Achieved fully.	Not achieved—the investment expenditures remain \$55.50 billion below their estimate.
P_8	Not achieved—the category-II government expenditures remained below their planned level by \$34.10 billion.	Not achieved—the planned government expenditure of category-II increased by \$3.8 billions, but still it was \$27.3 billion below its planned level.	Not achieved—the personal consumption expenditures remained \$28.3 billion below their limit.
P_9	Overachieved—personal and government savings plus the exports were higher by \$99.9 billions than the planned amount.	Not achieved—no change in the overachievement of \$99.9 billions in personal and government savings plus exports from run No. 1.	Not achieved—the personal and government savings plus the exports are overachieved by \$100.90 billion.

Run 2

The solution of this run clearly indicates that when the full employment goal is assigned top priority, then there is an increase in government purchase of goods and services (x_6 through x_{15}); on the other hand, the individual transfer payments, x_{16}, decrease significantly and the subsidies, x_{19}, are completely eliminated. This implies that the attainment of full employment and control of inflation are conflicting objectives. This is confirmed by the increased over-achievement of \$279.2 billion in

the economic stability, which means creation of a domestic growth gap. The net interest payments on the public debt are kept constant. This in turn may be interpreted as government's aversion to a decrease or increase in the level of public debt. Further, this run also indicates that the predicted spending level in the important areas of 'government expenditures-I' is first fully met, and the next spending level for not-so-important areas of 'government expenditures-II' also goes up by $6.8 billion (= 34.1 − 27.3) as compared to the corresponding expenditures-II of run No. 1.

Run 3

With the above-mentioned changes in priorities and capital–output ratio, the solution of run No. 3 indicated that the levels of attainment of goals 1 through 3 remained unchanged from those of run No. 1. However, goal 4, the growth in productive capacity was overachieved by $18.8 billion. In this solution both personal savings, x_4, and government surplus, x_{20}, became zero. It is an indication of the savings gap. In spite of these two zero savings, the growth investment expenditure (i.e. $a_2 \cdot a_3 \cdot b_{11}$ + capacity overachievement = 1 × 0.03 × 1090.2 + 18.8 = $51.5 billions) could still be fully met from the given corporate retained earnings, b_7, plus the capital consumption allowance, b_8, equal to: $b_7 + b_8$ = 51.5 + 112.0 = $163.5 billion. This is $18.8 billion more than the level of gross investment expenditures (calculated above to be equal to $144.7 billions) required to achieve a 3% growth in productive capacity above that of 1973.

With a capital–output ratio of 1:1, the growth investment expenditure determined in this run corresponded to a growth rate, a_3 of 4.7% i.e.

$$\frac{a_2 \cdot a_3 \cdot b_{11} + 18.8}{a_2 \cdot a_3 \cdot b_{11}} \times 3\%.$$

This is more than 50% higher than the assumed growth rate of 3%, in spite of a 22.0% decrease in the required gross investment expenditures ($x_{21} + x_{22} + x_{23}$), (i.e. from $210.0 billion with the capital–output ratio of 3:1 in the other two runs to (144.7 + overachievement of 18.8) = $163.5 billion with the capital–output ratio of 1:1 in the present run).

Because of an existing link between investment spending and personal spending through personal savings, any reduction in investment spending results in an equal reduction in personal savings required to finance such an investment. In consequence, less personal savings will be required, and more money will be available for personal consumption expenditures ($x_1 + x_2 + x_3$). In run 3, the personal expenditures on services, x_3 only decreased by $4.00 billion as compared to those of runs No. 1 and No. 2, although the gross personal expenditures still remained $28.7 billion below their estimated limit.

4. SUMMARY CONCLUSIONS AND SOME OBSERVATIONS

The goal programming model described above seems to fully satisfy its intended purpose—that of providing economic planners with a tool for identifying gaps. The appropriate test of utility of such a model depends upon its intended use. The test of a descripitive model is in its predictive ability, whereas the test of a normative model is in its ability to help the decision-maker to study his various combinations of goals. In either case, the requirement for acceptance of this model is not its perfection but its workability and efficiency compared to the next best available models—the gap models. Some of the results provided by this model may seem to be of less than desired practical value. However, this model does offer a framework, using which an economic planner can choose the most desirable mix of economic goals, instead of having just an intuitive feel or guesstimate of the outcomes from his chosen policies. This model can be further refined by bringing in several additional variables and several other constraints to identify additional gaps that may be important to an economic planner. The predictive capability of this model may additionally be increased by its users, as they become more and more familiar with all its economic relationships and subtleties of the involved assumptions. With growth of users' experience, even the accuracy of input data going into the model would increase and hopefully this model would then be in a position to identify all the important economic gaps that might be needing immediate attention. In particular, the three solutions of the model help us in making the following conclusions:

1. Through the use of a goal programming

model, it was predicted that the year 1974 was going to be a year of continuing inflation and unemployment. The model highlighted the existence of serious foreign exchange gap and domestic growth gap. These were the major concerns of the planners during 1974–75. In consequence, some major actions were taken and the growth gap has now been reduced considerably.

2. The results of the first run implied that high incomes cause demand to be in excess of available supply, thereby creating GNP gaps and inciting inflation. One possible course of action might be the imposition of wage and price controls. This may have the effect of restraining incomes from further increases and in consequence may stabilize demand. The programs aimed at closing GNP gap by way of increasing the productivity of both the labor and the capital resources may be implemented concurrently with wage and price freeze to achieve a goal of increasing total supply. As the supply and demand come more into line in real terms, the controls may then be phased out.

3. An alternative course of action seems to be to increase taxes on both the higher individual incomes and higher business profits. Concurrent with the higher taxes, a general reduction in government spending must also be achieved. These actions may eliminate savings gap and the surplus of government receipts over expenditures may then be used to subsidize research and development projects to find solutions to the increasing costs of energy, primary materials and decreasing productivity. Special credits may be given to the businesses which invest in such type of research and development projects. If no such positive course of action can be taken to put a halt to the rising inflation, then the nation must be willing to bear the cost of prolonged recession and unemployment [5].

4. Often it has been suggested that an aggregative planning model should contain measures of social goals such as pollution control, population control, increased productivity together with the economic goals as used in this model. Research into the area of goal programming and social planning has already been performed by other academic groups [3]. Their findings indicate that the inclusion of social goals in a model based on national accounts is fraught with difficulties. Expenditures for social programs are usually hidden within the aggregate investment expenditures of each individual year [1]. Since these expenditures cannot be easily separated, therefore there is no definite way to determine their optimum levels of spending.

REFERENCES

1. Bureau of Economic Analysis (1974) Pollution abatement expenditures and the national economic accounts. *Surv. Curr. Bus.* **54**(7), 58.
2. Bureau of Labor Statistics (1973) *Mon. Labor Rev.* **96**, 27–42.
3. Charnes A, Colantoni C, Cooper WW and Kortanek KO (1972) Economic, social and enterprise accounting and mathematical models. *Acc Rev.* **XLVII**(1), 98–99.
4. Economic Report of the President (1974) pp. 27, 43–47, 69, 249–250, 272. US Government Printing Office, Washington, DC.
5. Herendeen JB (1975) *Economics of the Corporate Economy*, Chap. 13. Dunnellen, New York.
6. Hertzberg MP, Jacobs AI and Trevathan JE (1974) The ultilization of manufacturing capacity, 1966–1973. *Surv. Curr. Bus.* **54**(7), 53.
7. Lee SM (1972) *Goal Programming for Decision Analysis.* Auerbach, Philadelphia.
8. McConnell and Campbell R (1972) *Economics: Principles, Problems and Policies*, 5th Ed., pp. 164–165, 181–182, 241–242, 357. McGraw Hill, New York.
9. Musgrave JC (1974) New estimates of fixed nonresidential business capital in the US, 1925–1973. *Surv. Curr. Bus.* **54**(3), 25.
10. Spulber N and Horowitz I (1976) *Quantitative Economic Planning and Policy*, pp. 158, 172–174, 237, 249, 253 and 343. Norton, New York.
11. Theory Deserts the Forecasters (1974) *Bus. Week* **2339**, 50.
12. Walker FV (1963) *Growth Employment and the Price Level*, pp. 66–71. Prentice Hall, Englewood Cliffs.

ADDRESS FOR CORRESPONDENCE: *Professor SC Aggarwal, College of Business Administration, The Pennsylvania State University, 609, Business Administration Building, University Park, Pennsylvania 16802, USA.*

APPENDIX

MODEL CONSTRAINTS*

$$x_1 + x_2 + x_3 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{21} + x_{22} + x_{23} + x_{25} - x_{26} + d_1^- - d_1^+ = 1125.2 = b_9 \quad (1)$$

$$0.6\,(x_1 + x_2 + x_3 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{21} + x_{22} + x_{23} + x_{25} - x_{26}) + d_2^- - d_2^+ = 884.1 = b_{10} \quad (2)$$

$$x_{21} + x_{22} + x_{23} + d_3^- - d_3^+ = 210.1 = a_2a_3b_{11} + b_8 \quad (3)$$

$$x_1 + x_2 + x_3 + x_4 + x_5 = 877.30 = b_1 - b_2 \quad (4)$$

$$x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{19} + x_{20} = 402.7 = b_3 + b_4 + b_5 + b_6 \quad (5)$$

Vol. 6
1978

$$x_{21} + x_{22} + x_{23} + x_{24} - x_4 - x_{20} = 163.50 = b_7 + b_8 \quad (6)$$

$$x_{25} - x_{26} - x_{24} - x_{17} - x_5 = 0.0 \quad (7)$$

$$x_1 + d_4^- = 145.9 \quad (8)$$

$$x_2 + d_5^- = 378.6 \quad (9)$$

$$x_3 + d_6^- = 377.1 \quad (10)$$

$$x_4 + d_7^- - d_7^+ = 74.4 \quad (11)$$

$$x_5 + d_8^- - d_8^+ = 1.21 \quad (12)$$

$$x_6 + d_9^- = 77.9 \quad (13)$$

$$x_7 + d_{10}^- = 2.3 \quad (14)$$

$$x_8 + d_{11}^- = 26.6 \quad (15)$$

$$x_9 + d_{12}^- = 0.1 \quad (16)$$

$$x_{10} + d_{13}^- = 75.9 \quad (17)$$

$$x_{11} + d_{14}^- = 56.8 \quad (18)$$

$$x_{12} + d_{15}^- = 3.5 \quad (19)$$

$$x_{13} + d_{16}^- = 34.9 \quad (20)$$

$$x_{14} + d_{17}^- = 1.6 \quad (21)$$

$$x_{15} + d_{18}^- = 7.8 \quad (22)$$

$$x_{16} + d_{19}^- = 83.8 \quad (23)$$

$$x_{17} + d_{20}^- = 25.7 \quad (24)$$

$$x_{18} = 14.4 \quad (25)$$

$$x_{19} + d_{21}^- = 10.1 \quad (26)$$

$$x_{20} - d_{22}^+ = 0.0 \quad (27)$$

$$x_{21} - d_{23}^+ = 48.6 \quad (28)$$

$$x_{22} - d_{24}^+ = 153.0 \quad (29)$$

$$x_{23} - d_{25}^+ = 17.4 \quad (30)$$

$$x_{24} + d_{26}^- - d_{26}^+ = 1.0 \quad (31)$$

$$x_{25} - d_{27}^+ = 138.6 \quad (32)$$

$$x_{26} + d_{28}^- = 138.4 \quad (33)$$

* All the right-hand side constants are expressed in billion $.

Vol. 6
1978

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 173–181
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0173/$02.00/0

An Empirical Application of Multidimensional Scaling to Advertising Pretesting

RE WILKES

Texas Tech University

EB UHR

Virginia Polytechnic Institute and State University

(*Received May* 1977; *in revised form September* 1977)

One of the least-managed aspects of advertising management may well be that of effectiveness measurement. The problem of obtaining satisfactory indications of advertising effectiveness is particularly critical in pretesting since the rationale for pretesting is to evaluate advertisements before their release for broadscale media distribution. Therefore, the primary purpose of the research reported here is to provide empirical evidence of the value of multidimensional scaling methods to pretesting advertising. The article demonstrates how nonmetric methodology obviates the disadvantages of such methods as consumer jury tests and rating scales. A secondary purpose of the research project is to demonstrate the value of pretesting itself by comparing the perceptions of target market consumers with those of selected group of advertising professionals.

ONE of the least-managed aspects of advertising management may well be that of effectiveness measurement. Whether advertising effectiveness can, in fact, be measured at all is a persistent question, although the majority of leading advertisers use some type of effectiveness measure [5]. And, while measurement techniques are becoming more sophisticated, doubts about effectiveness measurement methodology remain:

> The problem of copy-testing is not the theory, which is magnificent, but the practice, which is ludicrous. Advertising is a craft which has . . . most of the disadvantages of art [2].

> About 50% of the time a test result is likely to be inconclusive in telling us whether a given commercial is a 'good' one or a 'poor' one [11].

> It is strange that management is willing to devote funds to new-product development but is loathe to improve the tools with which advertising is measured [5].

The problem of obtaining satisfactory indications of advertising effectiveness is particularly critical in pretesting since the rationale for pretesting is to evaluate advertisements *before* their release for broadscale media distribution. A wrong decision at this 'go/no go' point may mean wasted production and media costs as well as a lowered probability of market success. Selection of the most appropriate pretest is complicated, however, by the variety of methods available and the fact that no one method, technique, or measurement variable is appropriate in all situations.

Ray [12] classifies pretests according to the stages of the advertising decision-making process: (1) developmental tests done as the advertising is being created; (2) selection-scheduling tests which are done with finished advertising and which seek to determine which of several alternatives should be run and how they should be scheduled; and (3) limited pretests such as the day-after-recall test used when the alternatives have been narrowed to one or a very few. The particular test(s) utilized inevitably trades off costs versus reliability and validity. For example, order-of-merit tests, while inexpensive and simple to administer, may only yield the best of the worst of a set of alternatives. Full-scale field experiments, however,

may require considerable financial outlay to attain high validity. According to Engel, the need for additional empirical research and improved methodology is apparent, with the need increasing more rapidly than it is being met [5, p. 361].

Paradoxically, however, the development of new methodology leads to claims and counterclaims as to which technique is suitable for which problem, or, even worse, the lack of application of otherwise appropriate methodology to present problems. For example, although Green and Carmone proposed in 1970 that multidimensional scaling methods might be profitably applied to advertising pretesting [6], the reported applications of this methodology have been primarily in other areas of advertising management, e.g. concept testing and positioning [15–17].

The primary purpose of the research reported here is to provide empirical evidence of the value of multidimensional scaling methods to pretesting advertising. The article demonstrates how nonmetric methodology obviates the disadvantages of such methods as consumer jury tests and rating scales (as presently utilized), while being highly compatible with other pretesting procedures the advertising strategist might also desire to employ. A secondary purpose of the research reported here is to demonstrate the value of pretesting itself by comparing the perceptions of target market consumers with those of a selected group of advertising professionals.

METHODOLOGY OF THE STUDY

Six different versions of a monochromatic advertisement for a well-known catalog seed company were made available by the client's advertising agency. The agency wanted to determine which of the six conceptualizations was the most effective for the client, that is, which was the 'best' advertisement for the campaign theme that had been selected. In terms of Ray's [12] classification, this procedure occupies the middle ground in pretesting; that is, the study involves the selection-scheduling type of pretest mentioned earlier. Three advantages of such tests are: (1) the use of finished advertisements to determine actual potential of ad response; (2) the laboratory setting, which offers the potential for control of those variables crucial to advertising effectiveness; and (3) selection-scheduling tests can be done quickly and with less cost than the limited post-testing type. And, in terms of Kahn and Light's [9] dichotomy of communication versus persuasion copy-testing systems, the test proposed in this paper falls in the communications domain.

Two groups of respondents provided judgments for the study. The first group consisted of 24 adult women selected to be representative of the client's target market according to known market profiles. In actual practice, larger samples might be desired for a better cross-section of the market segment. The small sample here is only for illustration of the scaling methodology. The second group was made up of 13 advertising professionals from the advertising club of a medium-sized city. The professionals were told the objective of the advertising campaign and also briefed as to the target market profile. Each group was first presented with all 15 distinct pairs of the six advertisements and instructed to rate each pair on the perceived dissimilarity using a scale from one (highly similar) to nine (highly dissimilar). This method of gathering proximity judgments has been shown to have identifiable advantages over other methods for gathering similarities data and yields unconditional data matrices for input to the multidimensional scaling algorithm [14].

Respondents next rated each advertisement on a seven-point intensity scale according to the perceived applicability of the following eight sets of bipolar adjectives:

exciting/dull
informative/unclear
trustworthy/suspicious
neat/cluttered
pleasant/unpleasant
good/bad
easty to read/hard to read
encourages my interest/discourages my interest

These criteria were selected for illustrative purposes only. Each advertiser would determine what criteria were most appropriate in his own situation. These criteria could be drawn from past experience, separate research, and other sources. These judgments yielded data matrices that were used to examine the way in which

Vol. 6
1978

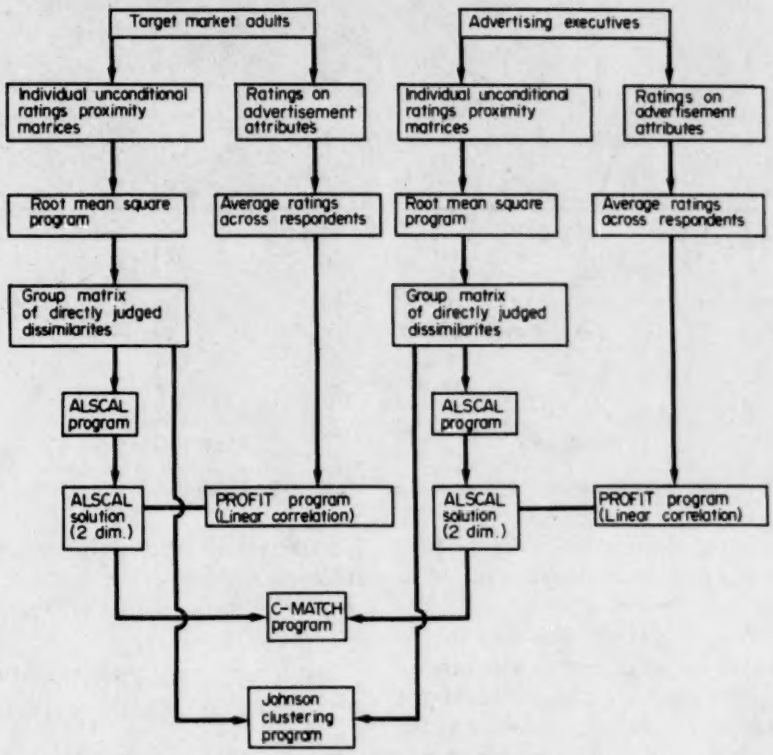


FIG. 1. *Flow diagram of analysis.*

the eight bipolar scale ratings related to the stimulus configurations obtained from the directly judged dissimilarities matrices. This analysis aids in the interpretation of the stimulus spaces obtained, as well as in the examination of any differences between two groups, as with the target consumers and advertising professionals in the present study.

Processing the data matrices

Figure 1 provides a flow diagram of the full analysis used in the study, following the very useful diagrammatic presentation of Green and Rao [7]. The directly judged dissimilarities were averaged across individuals using the root mean square rather than the arithmetic mean.[1] Young [18] has observed that if the dissimilarity judgements are in fact related to distances, then it is best to square the judgments before taking the mean since distances are not additive, whereas squared distances are. This is accomplished by use of the root mean square. The resultant matrix of dissimilarity judgments was then scaled in two dimensions for both groups of respondents through use of the ALSCAL algorithm [13]. Unlike most scaling routines that employ either a gradient procedure for fitting the Euclidean model or impose undesirable data restrictions, ALSCAL fits the model directly to the data (not to scalar products determined from the data) by an alternating least squares procedure which is convergent, nonoscillatory, quick, and relatively free from local minimum problems. A cluster analysis was performed on the dissimilarities data matrix as a check on the dimensionality of the configuration using Johnson's hierarchical cluster analysis procedure [8]. This method is invariant under monotone transformations of the input data. Correspondence between the configurations produced by scaling the directly judged dissimilarities matrices for the adult consumers and the advertising professionals was assessed through the use of Cliff's C-MATCH program [4]. This algorithm orthogonally rotates two matrices (e.g. two multidimensional scaling solutions) to congruence,

[1] An *n*-way replicated scaling of the data matrices was also executed because of the small number of stimuli in this set. The resultant configurations were virtually the same as from the averaged data matrices; stress values were in the same range as for the averaged data scalings.

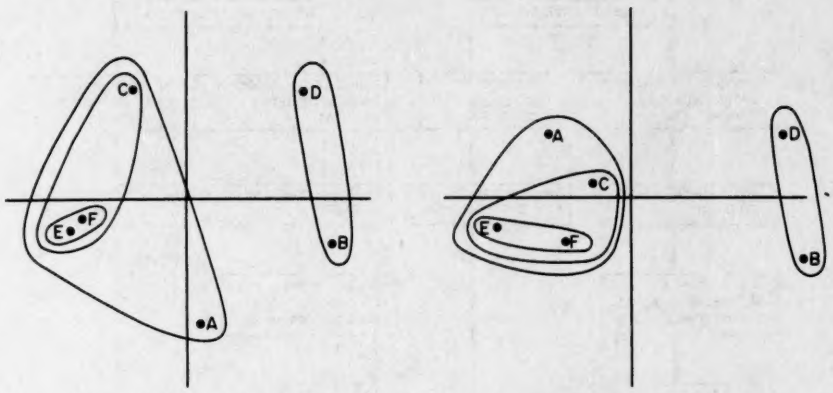


FIG. 2. *Two-dimension configuration from ALSCAL for both groups of respondents.*

either by rotating them simultaneously to a compromise position or by designating one as a target and rotating the other to maximum congruence to it. The latter case was utilized in this paper. Two measures of goodness of fit are computed: the average cosine of the angle between corresponding vectors and the product moment correlation of interpoint distances between target and data.

To determine how each group's bipolar scale ratings related to its stimulus configuration produced by the ALSCAL routine, Carroll and Chang's PORFIT algorithm [3] was utilized. PROFIT is a procedure for fitting outside product vectors into stimulus spaces. Such fitting involves finding, for each property vector separately (i.e. each bipolar scale or construct), a direction (vector) in the stimulus space whose stimulus projections are maximally correlated with the subject's property vector's scale values. Property fitting methods make possible a more intensive study of the differences in the configurations of two or more groups of respondents and also assist in axes labeling, although the latter use was not of primary concern in the study reported here.

RECOVERING THE UNDERLYING STRUCTURE OF THE DATA

The problem of multidimensional scaling may be broadly stated as that of finding n points whose interpoint distances match in some sense the experimental dissimilarities of n objects. That is, the researcher desires to uncover the pattern or structure in the data obtained and represent that structure as points in a spatial model in such a way that the significant features of the data about these objects are revealed in the geometrical relations among the points. Figure 2 presents the configurations produced by scaling the directly judged dissimilarities data matrices for both groups of respondents in the study reported here. Results from the clustering analysis are embedded within the configurations in the conventional method of plotted contours which successively incorporate increasingly dissimilar stimuli. The lack of distortion evident for these contours indicates the two-dimensional representation and is adequate. Additionally, the SSTRESS (1) values[2] are quite low, giving further evidence of the ability of the configurations to represent satisfactorily in two dimensions the underlying structure or pattern in the two data matrices; accompanying Shepard diagrams confirmed this judgment.

Looking first at Fig. 2a for the target consumers, it may be seen that advertisements E

[2] ALSCAL adopts a minimization criterion that is in the STRESS framework but differs due to being defined on squared distances and squared estimates; thus the formula is referred to as SSTRESS. While SSTRESS and STRESS are not strictly equivalent, the monotonic restriction

$$t^0[o^2_{ijk}] = [d^{*2}_{ijk}]$$

(where o denotes the squared observations, d^* the optimally scaled squared observations, i.e. the disparities, and t^0 the order restriction) defined on o^2_{ijk} and d^{*2}_{ijk} (i.e. the least squares estimate of d^2_{ijk}) is precisely equivalent to the monotonic restriction defined on o_{ijk} and d^*_{ijk}. There is no *a priori* reason for selecting one formula over the other, and ALSCAL selects SSTRESS for algorithmic convenience.

Vol. 6
1978

and F were perceived as highly similar, with C being the next most similar advertisement. At the opposite end of the *X*-axis (which itself represents one of the factors or dimensions respondents were using to make the dissimilarity judgements) are advertisements B and D. The other factor or dimension is represented by the *Y*-axis. Note that while B is perceived as quite different from E and F with respect to the first dimension, these three advertisements are seen as similar with respect to the second dimension. Advertisement A is seen as the opposite of advertisements C and D with respect to the *Y*-axis or second factor used in judging these dissimilarities. Both B and D are isolated from the remaining advertisements and are viewed as similar to each other. At this stage of the analysis, the interest is more in how the advertisements are spatially arranged than with actually labeling the axes. The researcher could order the six advertisements by projecting each one to the two axes and see if the bases for the ordering are discernible. The property fitting discussed later sheds light on this task; thus, no labeling is attempted here.

The advertising professionals (Fig. 2b) perceived advertisements B and D in a manner much the same as the target consumers. Also, both groups yielded dissimilarities judgments that place B and D opposite the remaining advertisements along the *X*-axis. Although the professionals judged E and F to be similar, they perceived less similarity between these two advertisements than did the target consumers. Further, advertisements E, F, C, and A are more tightly patterned in Fig. 2b than in Fig. 2a. There is also much less dispersion along the *Y*-axis in Fig. 2b, indicating that the second dimension (or judgment criterion) was less relevant to the executives than to the target consumers. It is also interesting to note the positioning of advertisement A in the two configurations. Whereas the adult women in the study perceived A and C at opposite ends of the *Y*-axis, with A seen as more similar to B than D, the professionals placed A closest to C and more similar, with respect to the *Y*-axis, to D than to B.

One procedure for determining how different these two configurations are is to set one configuration as a target matrix and rotating the other to congruence by maximizing their sum of cross-products. The two-dimension configuration obtained from scaling the dissimilarities of the target market consumers was held fixed and the configuration of the advertising professionals was fitted to it using Cliff's C-MATCH routine [4]. Both configurations were first translated to zero mean and rotated to principal axes. The best-fitting rotation of the 'data' matrix to the 'target' matrix involved a rotation of approx 13 degrees. The most appropriate measure for appraising goodness of fit, the product–moment correlation between interpoint distances [7], was only 0.594. Even if differences in vector length are ignored and use is made of the average cosine between pairs of vectors in the matched space, the goodness of fit increases only to 0.611.

At this stage in the analysis, several useful bits of information are provided. First, we have some understanding of the relative similarity of six proposed advertisements as perceived by a sample of target market consumers. We know that certain of these advertisements are highly similar and others quite dissimilar. Their arrangement in two-dimension space provides an interpretable configuration to the analyst. It may frequently be possible at this point to label the underlying dimensions (judgments criteria). We also know at this stage that advertising professionals display markedly different perceptions of these advertisements, not only with respect to each other but also with respect to the underlying criteria used for their perceptual judgments.

Objective property fitting

As stated earlier in this article, the researcher may facilitate the interpretation of multidimensional scaling solutions through the use of objective property fitting procedures. That is, we wish to know how the eight bipolar scales relate to the configuration produced by the ALSCAL program. After averaging across respondents in each group, ratings of the six advertisements on the eight bipolar scales were fitted to the corresponding configuration in Fig. 2 by use of the PROFIT routine [3]. This algorithm permits both a linear and nonlinear fitting to be selected. These two procedures yielded very similar results, however, and only the linear fitting is reported here. The linear fitting finds the best direction in the configuration space for each property (i.e. bipolar scale) such that the projections of the advertisements

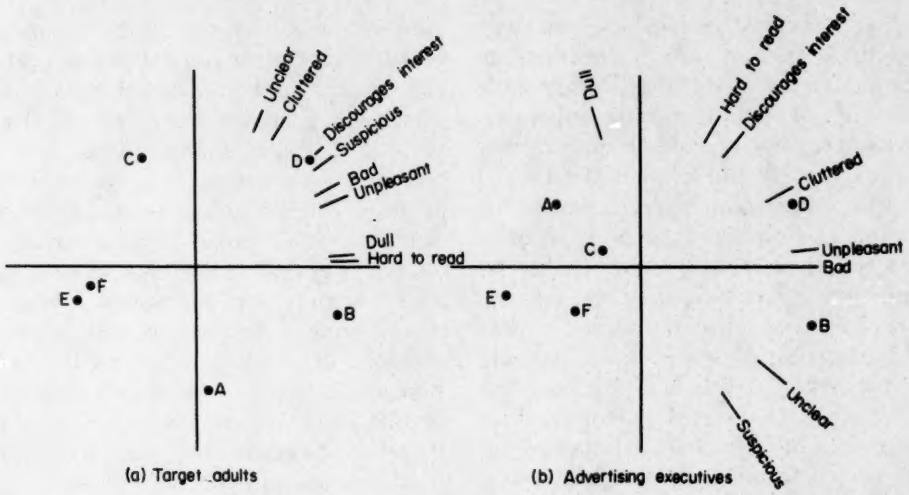


FIG. 3. *Eight property vectors fitted in ALSCAL configuration space for both groups of respondents using linear regression procedure.*

onto the fitted vector maximally correlate (linearly) with the original ratings on the property. This procedure essentially amounts to a multiple linear regression with each property scale serving in turn as a criterion variable.

Figures 3a and 3b portray the fitted vectors for the target consumers and the advertising professionals, respectively. The multiple correlation coefficients presented in Table 1 indicate the degree of fit between each fitted vector and the configuration for both groups of respondents. These signify how well each construct (represented by the property scale) is linearly related to the dimensions of the configuration.

It may be noted properties 1 and 2 (dull/exciting and unclear/informative) were not highly relevant to the target market sample; i.e. all had correlations under 0.6. Property 3 (suspicious/trustworthy), however, was viewed by this group as a very important construct (correlation = 0.99). Properties 5, 6, and 7 were also moderately important constructs to this particular group.

TABLE 1. MAXIMUM CORRELATION BETWEEN THE PROPERTY AND THE PROJECTIONS ON THE FITTED VECTOR FOR BOTH GROUPS OF RESPONDENTS

Property	Target adults	Advertising professionals
Exciting/dull	0.51	0.31
Informative/unclear	0.58	0.70
Trustworthy/suspicious	0.99	0.87
Neat/cluttered	0.64	0.79
Pleasant/unpleasant	0.80	0.83
Good/bad	0.80	0.62
Easy to read/hard to read	0.79	0.90
Encourages my interest/ discourages my interest	0.73	0.95

A useful way of looking at these property vectors is to rank order the advertisements on each construct, in turn, through perpendicular lines from each stimulus to each vector. This permits the analyst to determine how each advertisement 'performed' on the various constructs, which themselves may be ordered in perceived relevance according to the correlations in Table 1. Table 2 presents the rank orderings for the set of advertisements along each property vector for both groups of respondents. Advertisement E, followed by F, ranks first on the three constructs most highly correlated with the perceptual configuration for the target group. Advertisement D, followed by B, ranks last on these same constructs. Consequently, advertisements E and F would be considered to be the 'best' choices of the six alternatives presented, according to the perceptions of the target market sample.

These same constructs were perceived quite differently by the advertising professionals. Referring again to Table 1, property 8 (encourages/discourages interest) shows the highest correlation with the perceptual configuration (0.95), followed by property 7 (easy/hard to read) with a correlation of 0.90. Corresponding correlations for the target adults were 0.73 and 0.79, respectively. Further evidence of the perceptual differences between these two groups is clearly seen in Fig. 3b and the right half

TABLE 2. RANK ORDER OF SIX ADVERTISEMENTS ALONG EIGHT CONSTRUCTS FOR BOTH GROUPS OF RESPONDENTS

	Target adults Rank order						Advertising professionals Rank order					
Construct	1	2	3	4	5	6	1	2	3	4	5	6
Exciting	E	F	C	A	D	B	B	F	D	E	C	A
Informative	A	E	F	B	C	D	A	E	C	F	D	B
Trustworthy	E	F	A	C	B	D	B	D	F	C	E	A
Neat	A	E	F	B	C	D	E	F	A	C	B	D
Pleasant	E	F	A	C	B	D	E	A	F	C	D	B
Good	E	F	A	C	B	D	E	A	F	C	D	B
Easy to read	E	F	C	A	D	B	E	F	C	A	B	D
Encourages interest	E	F	A	C	B	D	E	F	C	A	B	D

of Table 2. Construct 3 (trustworthy/suspicious) is positioned in Fig. 3b such that advertisements B and D rank first and second, respectively, according to the advertising professionals, but last as perceived by the adult women. Recall that this construct had the highest correlation coefficient for the latter group. An important shift in ordering occurs for constructs 5 and 6, also of high relevance to the target consumer sample. The seemingly clear preference for advertisements E and F by the target group is not implicit in the property fitting results for the professionals on the three constructs the target group rated most important (i.e. 3, 5, and 6).

DISCUSSION

The central purpose of this article has been to demonstrate to advertising management the applicability and potential value of nonmetric methodology to advertisement pretesting. For the representative project utilized in this effort, advertising management specified that the pretesting problem to be solved was the selection of the 'best' alternative advertisement and the reasons for the particular selection. Essentially then, the advertising agency desired an approach that was both predictive and diagnostic.

In predicting which of a set of alternative advertisements should be pursued, care must be taken to insure that not only is the 'best' of the set selected, but that in addition, the advertisement performs well on salient objective criteria. The likelihood of selecting 'the best of the worst,' as might happen in portfolio analysis, for example, must be minimized. By fitting the objective property vectors to the multidimensional scaling space, positive and negative quadrants can be established; Fig. 3a shows that the negative quadrant, i.e. undesirable attributes, is the upper right portion of the space, while the lower left quadrant forms the positive portion or desirable attributes of the space. Clearly then, advertisements E and F are the superior executions in this set of advertisements. It is clear from Fig. 3a and Table 2 that advertisement E is the 'best' advertisement of the set created by the advertising agency since E is the most positive in Fig. 3a and ranks highest on the objective criteria in Table 2.

The reader will recognize that the above discussion utilized the results of the adult perceptions rather than those of the advertising professionals. Actually, in this case, the same advertisement would be selected as 'best' by both groups but for somewhat different reasons as discussed earlier in this article. What is important here is the clear indication that executive 'armchair' judgment cannot replace actual consumer perceptions. Even if the strength of the correlations presented in Table 1 are ignored and *only* the rank order of the properties is considered, two of the first three properties utilized by the advertising professionals *do not* coincide with those used by the intended market. Further, the one property coinciding is ranked third by the professionals while it is ranked first by the target adults. Clearly, advertising professionals should not be used in place of target market perceptions during pretest research.

How sensitive to different respondent groups is the approach illustrated here? Although the respondent sets used here were merely to demonstrate the usefulness of multidimensional

scaling techniques, the practitioner would be expected to use proper sampling procedures to assure external validity. Perceptual judgments gathered from divergent respondents would be as damaging in the use of nonmetric scaling techniques as in any measurement situation. In the study reported here, respondents from the target market were selected to be representative according to criteria provided by the client. A different group of advertising professionals may have yielded similar or substantially different judgments to those reported here. As this article sought more to demonstrate the various scaling procedures, this point was not given major consideration.

In addition to determining the 'best' advertisement, it is also desirable to determine the factors used in this determination so that improvements might be made and that an inventory of evaluative dimensions for future research projects can be established. This information is readily provided by the nonmetric methodology presented here. Critical attributes may be obtained from Table 1, and the direction of modification to existing advertisements on these attributes is given by Figs. 3a and 3b. Notice that even the 'best' advertisement, E, can be improved in terms of the two key attributes, trustworthiness and pleasantness, as well as the other attributes. This is indicated by E's position relatively near the origin in the multidimensional scaling output which demonstrates that the remainder of the positive quadrant, previously identified, is left for improvement. Additionally, advertising management should give particular attention to properties 5 and 6 rather than 7 and 8 since these properties are perceived to be more important by the target market. In future pretesting research, properties 2 and 4, of low importance to the target market, may be discarded in favor of new dimensions of concern. By updating the inventory of evaluative dimensions after each pretest, advertising management can build and maintain a set of consumer-based objective criteria.

Finally, it must be pointed out that the data requirements of the nonmetric methodology utilized in this investigation are principally the same as that collected during most pretesting research. The respondent is essentially requested to rank order a set of proposed advertisements and to evaluate each on a number of attributes. While data requirements are not increased, bias is avoided by securing the rank ordering before exposure to the evaluative criteria. In addition, relatively small samples, as used in order-of-merit of consumer jury tests, for example, are appropriate for the methodology specified. The analysis itself is computer based, utilizing 'canned' programs which minimize the amount of computer time to perform the analysis. Thus, the nonmetric methodology, while providing increased information and avoiding pitfalls inherent in more traditional pretest methods, does not increase the cost of advertising research. Properly used with other pretests to evaluate other aspects of commercial effectiveness such as persuasion and the effects of repetition, nonmetric methodology can be especially valuable to the pretesting problem.

REFERENCES

1. Aaker DA & Myers JG (1975) *Advertising Management*. Prentice-Hall, Englewood Cliffs.
2. Case G (1970) quoted in *J. Advertising Res.* **10** (22).
3. Chang JJ & Carroll JD (1969) *How to Use PROFIT—A Computer Program for Property Fitting, by Optimizing Nonlinear or Linear Correlation*. Bell Telephone Laboratories, Murray Hill, NJ (mimeographed).
4. Cliff N (1966) Orthogonal rotation to congruence. *Psychometrika* **31**, 33–42.
5. Engel JF, Wales HG & Warshaw MR (1975) *Promotional Strategy* (3rd Ed.). Irwin, Homewood, IL.
6. Green PE & Carmone FJ (1970) *Multidimensional Scaling and Related Techniques in Marketing Analysis*. Allyn & Bacon, Boston.
7. Green PE & Rao VR (1972) *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms*. Holt, Rinehart & Winston, New York.
8. Johnson SC (1967) Hierarchical clustering schemes. *Psychometrika* **32**, 241–254.
9. Kahn F & Light L (1972) Copytesting—communication vs. persuasion, in *Advances in Consumer Research*, Vol. 2. Proceedings of the 1974 Association for Consumer Research Chicago Conference.
10. Kotler P (1972) Editor's Foreword in *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms*. Holt, Rinehart & Winston, New York.
11. Longman KA (1971) *Advertising*. Harcourt Brace Jovanovich, New York.
12. Ray ML (1974) The advertising pretest as part of a multimeasure, multimethod, multisituation validation and application research system, in *Advances in Consumer Research*, Vol. 2. Proceedings of the 1974 Association for Consumer Research Chicago Conference.
13. Takane Y, Young FW & De Leeuw J (1975) Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. Report No. 147, The LL Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, NC.
14. Taylor JR (1969) Alternative methods for collecting similarities data, in *Marketing Involvement in Society*

and the Economy. American Marketing Association Proceedings of the Fall Conference.
15. WILKES RE Product Positioning: Confirmation via Multidimensional Scaling. *J. Advertising Res.* in preparation.
16. WIND Y (1973) A new procedure for concept evaluation. *J. Marketing* **37**, 2–11.
17. WIND Y & ROBINSON PJ (1972) Product positioning: an application of multidimensional scaling. *Attitude Research in Transition* (Ed. HALEY RI). American Marketing Association, Chicago.
18. YOUNG FW (1973) Conjoint Scaling. Report No. 118. The LL Thurstone Psychometric Laboratory. University of North Carolina, Chapel Hill, NC.

ADDRESS FOR CORRESPONDENCE: *Dr EB Uhr, Director of Graduate Programs in Business Administration, College of Business, Virginia Polytechnic and State University, Blacksburg VA* 24061, *USA.*

ol. 6
978

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 183–187
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0183 $02.00/0

Scheduling a Multi-Stage Production System with Startup Delays

JM MOGG

PE BIGHAM

The University of Houston

(Received January 1977; in revised form November 1977)

The problem treated is that of multi-stage scheduling as affected by the interdependence of successive stages. An algorithm is developed which incorporates startup delays as analytical functions of cycle times and production rates at adjacent stages. This is accomplished in part by imposing integer multiple restrictions on the relationship between the cycle times at adjacent stages. Total cost is then minimized for a given set of integer multiples.

Vol. 6
1978

SCHEDULING A MULTI-STAGE SYSTEM WITH STARTUP DELAYS

THE SYSTEM considered here is one in which N production stages are arranged in series, with the output of one stage serving as input to the next. Significant assumptions include deterministic parameters, an infinite time horizon, constant lot sizes, and finite production rates. The major obstacle to economic scheduling in this system is the interdependence of successive stages. This interdependence is accounted for by computing the delay in the initial startup at each stage, relative to that stages' predecessor, required to prevent interstage stockouts.

Although several of the listed references discuss multi-stage scheduling under these assumptions, most add restrictions which make startup delays unnecessary [2, 6–8]. Jensen and Khan [5] do not assume away the stage interdependence problem; however, their algorithm involves only the selection of the best schedule from among a user supplied list of possible schedules. Startup delays are found in [5] as non-analytical functions of the cycle times and production rates at adjacent stages.

In this paper, an integer multiple restriction is imposed on the relationship between the cycle times at adjacent stages. Analytical startup delay expressions are written, and total cost is minimized for a given set of integer multiples. A computerized algorithm for finding the optimal integer multiples values is then discussed.

MODEL DESCRIPTION

The multi-stage production model developed below has the following characteristics:

(1) A single product is produced for which demand occurs constantly and continuously.
(2) The production stages are arranged in series with raw materials at the input to the system and product units moving sequentially from one stage to the next until arriving at the output stage as the final product.
(3) Each production stage has a fixed known finite rate of production which is greater than the demand rate for the product. Thus, each stage must be periodically shut down and restarted so that the average production rate of the stage is equal to the demand rate. A setup cost is realized for each such operation. In addition, the uneven production of each stage causes in-process inventories to accumulate between the various stages with an associated inventory holding cost incurred as a result.

The following notation is used:

N = number of production stages (numbered in reverse order to the product flow).
D = demand (units/day) for the finished product.
R_i = production rate (units/day) of stage i while operating (all quantities are measured in terms of units of final product).
h_i = cost (\$/unit-day) of holding one unit of inventory for 1 day at stage i ($i = 2,\ldots,N$).
S_i = setup cost (\$) for stage i. This is the sum of the costs of one startup and one shutdown operation at stage i.
T_i = cycle time (days) for stage i, (i.e. the length of time between two successive startups at stage i).
t_i = production on-time (days) per cycle at stage i. Each stage must produce in each cycle the quantity required to meet the product demanded during the cycle. Therefore, equating total production per cycle at the ith stage (t_iR_i) to demand per cycle (T_iD), and solving for t_i yields

$$t_i = T_iD/R_i. \tag{1}$$

d_i = the delay (days) in the initial production startup at stage i, relative to that at stage $i + 1$, required to insure that no shortages occur in the in-process inventories held between the stages.

The scheduling methods presented here have as an objective the determination of the cyclic startup and shutdown schedules for each stage (i.e. the schedule of production lot sizes) which minimize total time-averaged cost. The only costs considered are the startup and shutdown costs incurred at each production stage, and the cost of holding the in-process and finished goods inventories. No inventory shortages are allowed, and an infinite time horizon and continuous review reorder policy are assumed. The setup and shutdown times required at each stage are initially assumed to be zero. (It can be shown that this assumption is not crucial.)

In addition to the above, a modified version of the integer multiple restriction imposed by others [2, 6–8] is adopted. In particular, the only production schedules considered are those for which

$$T_{i+1} = m_iT_i \tag{2}$$

where m_i is a positive integer or the reciprocal of a positive integer.

THE STARTUP-DELAY PROBLEM

Let the inventory held between stage i and its immediate predecessor, stage $i + 1$, be designated the stage $i + 1$ installation inventory [1]. The inventory profile associated with these units will cycle over a period of time equal to the least common multiple of the production cycle times at stages i and $i + 1$. This inventory profile will, in general, not have the familiar saw-tooth shape.

To illustrate, consider the parameter values and cycle times listed in Table 1. The stage $i + 1$ installation inventory profile resulting from these values is plotted in Fig. 1, Part A. As the graph shows, this inventory is negative over part of each cycle. This is, of course, a physical impossibility since stage i can operate only on the semi-finished units passed to it by stage $i + 1$.

The negative inventory positions illustrated in Fig. 1 may be avoided by delaying the initial production startup at stage i, relative to that

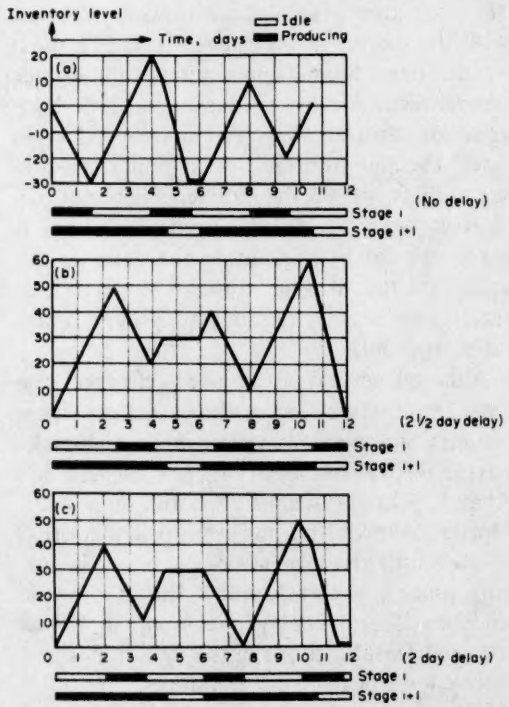


FIG. 1. *Stage $i + 1$ inventory profile for startup delays of 0, 2½, 2 days (reference Table 1).*

TABLE 1. PARAMETER VALUES OF HYPOTHETICAL STAGES. $R_i > R_{i+1}$: (TIMES IN DAYS, D = 15 PER day)

Stage j	Cycle time T_j	Production rate R_j	Production on time t_j	Production off time $T_j - t_j$
i	4.0	40.0	1.5	2.5
$i+1$	6.0	20.0	4.5	1.5

at stage $i + 1$. For example, suppose the initial start of production at stage i is delayed 2.5 days. The resulting inventory profile is shown in Fig. 1, Part B. Numerical integration techniques may now be used to calculate an average daily stage $i + 1$ inventory of 30 units. However, as Fig. 1, Part C shows, stage $i + 1$ stockouts could also have been avoided by a delay of only 2.0 days, with a resulting average daily inventory of 22.5 units. If stage $i + 1$ has the greater production rate, a startup delay at stage i may or may not be required, depending on the cycle times chosen.

In general, the minimum required delay at any stage i is a function of the production rates and cycle times at stages i and $i + 1$ (the preceding stage); and, as illustrated above, the length of this delay directly affects the size of the time-averaged inventory held between stages i and $i + 1$. Systematic methods for finding minimum required delay times using what may be termed 'deterministic simulation' have been developed [5]. However, no general closed-form (i.e. analytical) expressions have been derived for such delays.

Clark and Scarf [1] define the echelon of any stage i in a multi-stage system as the units in the system which have passed through that stage but which have not yet been sold. Under this concept, the inventory holding cost chargeable to stage i is the incremental holding cost incurred as a result of processing at that stage.

The echelon inventory of any stage i will be depleted constantly and continuously by final demand, and replenished by the cyclic production runs at stage i. Thus, if the effects of startup delays are ignored, the average number of unit-days held per day in this inventory is given by the single stage result

$$I_i = T_i D(R_i - D)/2R_i. \quad (3)$$

The total average setup and holding cost incurred at all N stages may then be written as

$$\Sigma_{i=1}^{N-1}[(S_i/T_i) + I_i h_i] + D\Sigma_{i=1}^{N-1}\Sigma_{j=i+1}^{N} h_j d_i \quad (4)$$

where the last term of (4) accounts for the total additional holding cost resulting from startup delays at stages 1 through $N - 1$. Minimization of (4) with respect to the T_i requires that the d_i be written as functions of these decision variables.

MINIMUM REQUIRED STARTUP DELAY TIME

Assume that for any two adjacent stages in the system described above there exists prime integers K_i and K_{i+1} such that

$$L_{i+1} = K_{i+1}T_{i+1} = K_i T_i. \quad (5)$$

where L_{i+1} is the least common multiple of T_i and T_{i+1}, and is the period of time over which the stage $i + 1$ installation inventory will cycle. Given the production rates at stages i and $i + 1$, general expressions for the minimum required startup delay at stage i are defined as below.

If $R_i \geq R_{i+1}$ the points of relative minimum for the stage $i + 1$ installation inventory occur whenever stage i shuts down. The minimum required delay at stage i is found to be

$$d_i = \max_k\{(kT_iD/R_{i+1}) + [kT_i/T_{i+1}](T_{i+1} - t_{i+1}) - ((k-1)T_i - t_i)\}; \quad (6)$$

where the first two terms in (6) are the point in time at which cumulative production at stage $i + 1$ equals cumulative production at stage i and the last term is the point in time at which i shuts down for the kth time. Also $[a]$ denotes the greatest integer strictly less than a.

If $R_i < R_{i+1}$, the points of relative minimum for the stage $i + 1$ installation inventory occur at those points in time at which stage $i + 1$ starts production. The minimum required delay at stage i is found to be

$$d_i = \max\{(k-1)T_{i-1} - ((k-1)T_{i-1}D/R_i) - [(k-1)T_{i-1}/T_i](T_i - t_i)\}; \quad (7)$$

where the first term in (7) is the point in time at which stage $i + 1$ starts production for the kth time and the last two terms the point in time at which cumulative production at stage i last equaled that at stage $i + 1$. Again, $[a]$ denotes the greatest integer strictly less than

TABLE 2. ANALYSIS OF HYPOTHETICAL 3 STAGE SYSTEM, $D = 15$ PER day

Stage (i)	System description S_i	R_i	H_i	Computed values T_i	d_i	Average daily costs set-up	hold	delay	Total cost
1	$100	40/ Day	$2	3.02	1.14	$33.03	$7.09	$25.54	$65.67
2	200	20	1.50	6.05	0	33.83	5.68	0	38.71
3	100	600	1	6.05	0	16.52	44.27	0	60.79
				Totals		82.59	57.04	25.54	165.71

a. The further development of analytical expressions for d_i must consider the integer multiple assumption defined by (2) and the magnitude of T_i *vis a vis* T_{i+1}. Similar to (2) the cycle time at any stage i may be written as a function of the cycle time at stage 1, namely $T_i = M_i T_1$, to be substituted into (4), which then becomes a strictly convex function of T_1. The optimum T_1 to minimize cost may then be determined.

AN APPLICATION

A computerized algorithm has been developed for finding the minimum cost set of integer multiples. Suppose that m_1' is the smallest positive integer for which

$$C(m_1') \leq C(m_1' + 1.0).$$

It may then be shown that m_1' is the optimal value for m_1, given the current values for m_i, and the restriction that m_1 be a positive integer. Similarly, if m_1'' is the smallest positive integer for which

$$C(1.0/m_1'') \leq C(1.0/(m_1'' + 1.0)),$$

the optimal value for m_1 is then $(1.0/m_1'')$, given the current values for m_i, and the restriction that m_1 be the reciprocal of a positive integer. The program written uses an iterative procedure to find $C(m_1')$ and $C(1.0/m_1'')$ for each possible set of integer multiples m_i, $i = 2,\ldots,N-1$. The least cost solution found to date is saved after each interation.

To illustrate, consider the 3-stage system described by the parameter values listed in Table 2. Use of these values yields unrestricted cycle times of 3.05, 6.53, and 5.30 days for stages 1, 2, and 3 respectively. Given (2), these cycle times imply $m_1 = 2.0$ and $m_2 = 1.0$, which yield the integer multiple production schedule described in Table 2.

Note that solution time for the above procedures expands exponentially in the number of stages considered. The authors have developed a heuristic search technique for finding optimal integer multiples. This method involves setting $m_i = 1.0$ and then making the single incrementation or decrementation which yields the largest cost savings. This procedure is repeated as long as cost savings may be attained. Fifty randomly generated problems were solved by both techniques; enumeration yielded a total cost more than 2% lower than that found by search in only three cases.

SUMMARY AND CONCLUDING REMARKS

The multi-stage system considered herein is essentially identical to that described in references [2] and [5]. In [2], integer multiple production schedules are found only for the case where $R_{i+1} \geq R_i$, thus avoiding the necessity of startup delays. In the present work, explicit expressions have been written for the d_i thereby allowing the removal of this restriction. Jensen and Khan [5] do not impose the integer multiple restriction, requiring instead that the user input finite lists of allowable cycle times for each stage. Their algorithm involves only the recursive selection of the least cost set of cycle times from among these user-defined possible sets. The performance of this algorithm is directly limited by the judiciousness with which possible cycle times are specified. No guidelines are developed in [5] for selection of allowable cycle times.

REFERENCES

1. CLARK A & SCARF H (1960) Optimal policies for a multi-echelon inventory problem. *Mgmt Sci.* **6**, 475–490.
2. CROWSTON WB, WAGNER M & WILLIAMS J (1973) Economic lot size determination in multi-stage assembly systems. *Mgmt Sci.* **19**(5), 517–527.

3. CULLENDER M (1971) A scheduling algorithm for a multi-stage multi-product production system. Unpublished masters thesis, The University of Texas.
4. HADLEY G & WHITTIN TM (1963) *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs, NJ.
5. JENSEN PA & KHAN HA (1972) Scheduling in a multi-stage production system with setup and inventory costs. *AIIE Trans.* **4**(2).
6. SCHUSSEL G (1968) Job shop lot release sizes. *Mgmt Sci.* **14**(8), B449–B472.
7. TAHA HA & SKEITH RW (1970) The economic lot sizes in multi-stage production systems. *AIIE Trans.* **2**(2).
8. THOMAS AB (1963) Optimizing a multi-stage production process. *Op. Res. Q.* **14**(2), 201–213.

ADDRESS FOR CORRESPONDENCE: *Dr JM Mogg, Associate Professor, University of Houston, Quantitative Management Science, Houston, Texas 77004, USA.*

Vol. 6
1978

BLANK
PAGE

Vol.
197

Vol. 6
1978

MEMORANDA

OME. 6/2—G

BLANK
PAGE

Vol.
197

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 191–192
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0191/$02.00/0

Vol. 6
1978

GERT Models of Discrete Probability Distributions and Sums

GERT (Graphical Evaluation and Review Technique) is a network modeling technique developed to analyze generalized stochastic networks [1–5]. The GERT simulation programs provide for computer simulation of models that can be depicted in network form. This paper demonstrates how to use the GERT technique to develop discrete probability distributions and the distribution of the sum of several independent random variables.

Consider first the problem of simulating the number of successes, x, obtained in n trials from a Bernoulli process. Observations on the variate X which is described by the binomial probability mass function

$$P(X = x|n, p) = \binom{n}{x} p^x q^{n-x} \text{ for } x = 0, 1, 2, \ldots, n \quad 0 < p < 1 \qquad (1)$$

can be generated using the GERT network of Fig. 1.

Nodes 2 and 5 of the network are the start and sink nodes, respectively. To generate one simulation or, equivalently, to generate the n Bernoulli trials, the number of network completions required for the realization of the sink node is set equal to n. Further examination of the model shows that node 3, which branches probabilistically to success or failure, represents the actual Bernoulli trial. Regardless of the outcome of any given trial, node 4 is realized and activities 4–5 and 4–3 are initiated. The completion of activity 4–5 corresponds to one of the n network completions required for the realization of the sink node. If, following the completion of activity 4–5, the number of network completions is less than n, activity 4–3 initiates the next Bernoulli trial.

Two important components of the above model are the F-statistic of the sink node and the activity times assigned to the network activities. The function of the F-statistic is to record the total network activity time required for the first realization of the corresponding node. Since the 'success' branch of node 3 has an activity time of one and the times assigned to all other network activities are zero, the F-statistic of node 5 acts as a counter and the total activity time recorded therefore represents the number of successes that occur in n trials of a Bernoulli process. After m simulations, the F-statistic provides: (1) the minimum and maximum number of success that occur; (2) the expected value and standard deviation of the number of successes, and (3) a histogram for the number of successes all related to n trials of the given Bernoulli process.

Like the binomial distribution, the negative binomial distribution concerns experiments that involve a Bernoulli process. Rather than considering the number of successes in n trials, the negative binomial random variate is defined as the number of Bernoulli trials required to obtain r successes. A GERT network that represents the negative binomial probability mass function

$$P(X = x|r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \text{ for } x = r, r+1, \ldots \quad 0 < p < 1 \quad r = 1, 2, \ldots \qquad (2)$$

is shown in Fig. 2.

Inspection of this model shows that the number of network completions necessary for the realization of node 5 corresponds to the parameter r in the distribution function. As in the network for the binomial distribution, a probabilistic node (node 3) that branches to a success or failure with probabilities p and 1-p, respectively, is utilized to represent the Bernoulli trial. If a simulated trial results in a success, node 4 is realized and activities 4–5 and 4–3 are initiated. Completion of the former activity represents one of the r network completions required for the realization of the sink node. If the number of completions is less than r, activity 4–3 initiates the next Bernoulli trial. The feedback

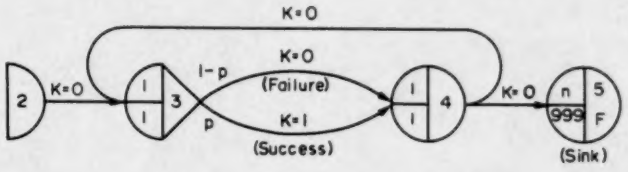


FIG. 1. *GERT network for the binomial distribution.*

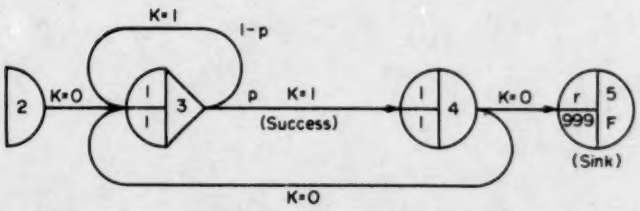


FIG. 2. *GERT network for the negative binomial distribution.*

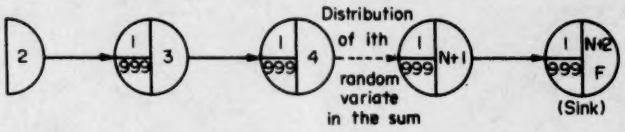


FIG. 3. *GERT network for the distribution of a sum of N independent random variables.*

loop, activity 3–3, serves the same function as activity 4–3 when the outcome of the Bernoulli trial is a failure.

Further examination of the network shows that the *F*-statistic of node 5 is actually a counter. Since both activity times assigned to the probabilistic branches of node 3 are one and all other activity times are zero, each time a Bernoulli trial is simulated the total network activity time is increased by one. Consequently, when the sink node is realized, the total network activity time represents the number of Bernoulli trials required to obtain the r successes. After m simulations of the experiment, the *F*-statistic provides: (1) the minimum and maximum number of trials required for the r successes, (2) the expected value and variance of the number of trials required for the r successes, and (3) a histogram for the number of trials required for the r successes.

The networks considered in the previous sections illustrate a few of the applications of GERT in the simulation of known discrete probability distributions. Consider now the somewhat more complicated problem of constructing generators for the distribution of a sum of independent random variables. A general model for this problem is shown in Fig. 3.

It can be seen that the GERT analog for the sum is simply a series of activities and deterministic nodes. The completion time for each activity represents a random variable and the distribution assigned to the activity time corresponds to the distribution of the random variable. Since the activities are independent, the total activity time recorded by the sink node *F*-statistic represents one observation on the sum of the N independent random variables.

The most current version of GERT [4] contains generators for ten probability functions and a user written function. If distributions of random variables in the sum are not among these, they can often be simulated using the technique previously outlined.

The fundamental basis for the simulation of stochastic systems is the generation of random variables that describe the various system activities. In this paper, GERT models for the simulation of several probability functions were presented to illustrate the application of the GERT in this area. The advantages inherent in this approach to the construction of pseudo-random number generators are twofold: (1) the conceptual simplicity of the GERT models and (2) the relevant statistics that are automatically provided by the GERT simulation.

REFERENCES

1. CLAYTON ER and MOORE LJ (1976) *GERT Modeling and Simulation: Fundamentals and Applications.* Petrocelli/Charter, New York.
2. PRITSKER AAB (1966) *GERT: Graphical Evaluation and Review Technique.* Rand Corp., Santa Monica, CA.
3. PRITSKER AAB and HAPP WW (1966) GERT: Graphical evaluation and review technique, Part I—fundamentals. *J. Engng* **17**(5), 267–274.
4. PRITSKER AAB (1977) *Modeling and Analysis Using Q-GERT Networks.* John Wiley, New York.
5. WHITEHOUSE GE and PRITSKER AAB (1969) GERT: part III—further statistical results; counters, renewal times and correlations. *AIIE Trans.* **1**(1), 45–50.

ER Clayton
EC Houck

(*September* 1977)

Department of Business Administration
College of Business
Virginia Polytechnic and State University
Blacksburg
VA 24061
USA

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 193–194
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0193/$02.00/0

Vol. 6
1978

Some Ranking Formulae for Alternative Research Projects

THE TERM 'alternative research projects' was used in [2] to denote a collection of projects with the property that success in research occurred at the first successful completion of one of the projects from the collection. Such projects were to be thought of as alternative possible routes to the same destination or research objective. The example was cited of new-product chemical research where success in research takes place when one of a collection of different formulations which have been suggested for some new product is tested successfully. Glazebrook [2] began an investigation of the problem of how research effort should be applied to a collection of alternative research projects in such a way that research success is achieved (i.e. one of the projects is completed successfully) as cheaply as possible. This memorandum is a report of further work in this area.

Cranston's [1] very simple model of a research project was adopted in [2]. Each project was assumed to consist of a number of research phases to be attempted in a pre-assigned order, each phase having a known probability of success and a known fixed cost. With this model the optimal strategy for the allocation of research effort was determined by a project ranking index. The theoretical basis for this result was provided by a little known result from the sequential design of experiments, called the Dynamic Allocation Index (DAI) Theorem. This result (explained in detail in [2]) states that a wide range of allocation problems have optimal strategies determined by ranking indices or DAI's.

This memorandum is a report of two new results. Optimal allocation strategies are given for problems incorporating more general project models than the one described above; namely, a model when costs are discounted instead of fixed, and a model in which the assumption that all phases within a project must be attempted in a pre-assigned order is dropped and a general precedence structure on the phases is allowed. This latter model corresponds to a general network representation for each project. The optimal strategy for this wide collection of problems has an intuitively appealing form.

DISCOUNTED COSTS

Assume that we have m alternative research opportunities or projects. Project i has n_i constituent phases, the j being denoted by (i, j), $j = 1(1)n_i$, $i = 1(1)m$. A single key resource is assumed to be available but may not be allocated to phase $(i, j+1)$ until (i, j) has been completed successfully. A research-phase once initiated must be completed but any project may be abandoned after the completion of a phase and may later be resumed from the point of abandonment without incurring any additional cost. Henceforth, an arbitrary research phase will be denoted by r.

Phase r is assumed to take X_r time units to be completed. If r is completed at time F_r (flow-time) a positive cost $C_r a^{F_r}$ is incurred where $0 \leq a < 1$. For convenience we write $M_r = a^{X_r}$. The probability of r being completed successfully is P_r. Success in research occurs as soon as one of the phases (i, n_i) is completed successfully. A success in research achieved in project i at time S_i earns a positive benefit (a negative cost) $B_i a^{S_i}$. Note that all the results in this section go through with very little modification if the $X_{(i,j)}$, $\forall_{j,i}$, are assumed to be independent random variables.

The strategy of allocating research effort to the projects which minimises the expected cost incurred up to research success is derived from the DAI theorem [2]. In the optimal strategy each project has associated with it throughout the evolution of the system an index called its DAI. This will change in value for a project as its phases are completed.

The DAI for a typical project with discounted costs is computed as follows. Suppose a project p initially consisted of phases $1, 2, \ldots, n$ (to be attempted in that order) but that of those $1, 2, \ldots, s-1$ have already been completed successfully. Before defining the DAI we need to introduce some other quantities.

Let $A[s, s+1, \ldots, t]$ be the expected cost incurred in attempting the phases $s, s+1, \ldots, t$ in order until any failure occurs, and $B[s, s+1, \ldots, t]$, roughly speaking, a discounted version of the expected time taken to complete the phases $s, s+1, \ldots, t$ or until a failure occurs among them.

i.e.

$$A[s, s+1, \ldots, t] = \begin{cases} \sum_{k=s}^{t} \left(\prod_{j=s}^{k-1} P_j M_j \right) C_k M_k, & 1 \leq s \leq t < n, \\ \sum_{k=s}^{n} \left(\prod_{j=s}^{k-1} P_j M_j \right) C_k M_k - \left(\prod_{j=s}^{n} P_j M_j \right) B_p, & 1 \leq s \leq t = n. \end{cases}$$

and

$$B[s, s+1, \ldots, t] = \begin{cases} \sum_{k=s}^{t} \left(\prod_{j=s}^{k-1} P_j M_j \right)(1 - M_k), & 1 \leq s \leq t < n, \\ \sum_{k=s}^{n} \left(\prod_{j=s}^{k-1} P_j M_j \right)(1 - M_k) + \prod_{k=s}^{n} P_k M_k, & 1 \leq s \leq t = n. \end{cases}$$

The DAI for project p after phases $1, 2, \ldots, s-1$ have been completed successfully is then $\gamma_p(s)$, given by

$$\gamma_p(s) = \min_{s \le t \le n} \{A[s, s+1, \ldots, t]/B[s, s+1, \ldots, t]\}. \quad (1)$$

We can summarise the above results, i.e.

Theorem 1

Suppose that at some time $s_i - 1$ of the n_i phases in project i have been completed, $i = 1(1)m$. An optimal strategy is always to allocate research effort to the next phase of any project which has not failed and which has the smallest associated value of $\gamma_i(s_i)$. If more than one project satisfies these criteria, it is immaterial which of them is chosen next. No more effort is allocated after success is achieved.

Theorem 1 follows from the DAI theorem in similar fashion to Theorem G in [2], though the proof is considerably more difficult technically. Note that, if a is allowed to tend to 1, $\gamma_i(s)$ converges to the fixed costs ranking index derived in [2], as might be expected.

General precedence structure

We consider here the problem discussed in the previous section with the following two modifications:

(1) a, the discount rate, is taken to be 1 (costs are fixed), and

(2) the phases in project i may be allocated research effort according to a general precedence relation R_i. We take $\{(i, j), (i, k)\} \epsilon R_i$ to mean that any valid allocation strategy must have the property that the phase (i, j) must be completed successfully before effort may be allocated to (i, k). It is required of R_i

(a) that it be *transitive*, that is $\{(i,j_1), (i,j_2)\}$, $\{(i,j_2), (i,j_3)\} \epsilon R_i$ imply that $\{(i, j_1)(i, j_3)\} \epsilon R_i$, and

(b) that there is at least one ordering of the phases within project i which is consistent with R_i (any such ordering is said to be *feasible*).

The set of feasible orderings for project i is denoted by F_i, $i = 1(1)m$. If $\alpha_i \epsilon F_i$ we denote by $C^i(\alpha_i)$ the expected cost which accrues from project i if it is allocated effort continuously according to α_i until a phase is unsuccessful. Define α_i^* to be *cost minimal* (not necessarily unique) for project i, i.e.

$$C^i(\alpha_i^*) = \min_{\alpha_i \epsilon F_i} \{C^i(\alpha_i)\}.$$

We shall assume without loss of generality that the ordering $\{(i, 1), (i, 2), \ldots, (i, n_i)\} \equiv \alpha_i^*$. The following theorem may then be proved:

Theorem 2

An optimal allocation strategy for the project selection problem with general precedence structure is to allocate research effort *within* each project according to α_i^* and *between* the projects as follows: if, at some time, $\{(i, j), j = 1(1)s_i - 1\}$ only of the phases within project i have been completed successfully, for $i = 1(1)m$, next allocate research effort to project which has not failed and which has the smallest associated value of

$$\left\{\sum_{k=s}^{n_i}\left(\prod_{r=s_i}^{k-1} P_{i,r}\right)C_{i,k} - \left(\prod_{r=s_i}^{n_i} P_{i,r}\right)B_i\right\}\left\{\prod_{r=s_i}^{n_i} P_{i,r}^{-1}\right\},$$

where $C_{i,j}$ and $P_{i,j}$ are the cost and probability of success respectively for (i, j).

If more than one project satisfies these criteria, it is immaterial which of them is chosen next. No more effort is allocated after success is achieved.

The proof of Theorem 2 and some results relating to the computation of cost minimal orderings may be found in [3]. Theorem 2 is not a corollary of the DAI theorem and the proof is by induction on the number of phases in all of the research projects. The reader should note that in Theorem 2 once it has been established that the within project strategy is cost minimal, the between project strategy follows from Theorem G in [2] since then the cost minimal ordering may be thought of as a pre-assigned ordering for the phases within each project. There is an approach to the computation of cost minimal orderings which is based on DAI's and which is explained in [3].

REFERENCES

1. Cranston RW (1975) First experiences with a ranking method for portfolio selection. *R&D Mgmt* **5**, 88–92.
2. Glazebrook KD (1976) A Profitability index for alternative research projects. *Omega* **4**(1), 79–83.
3. Glazebrook KD (1976) Stochastic scheduling. Ph.D. Thesis, University of Cambridge, UK.

Dr KD Glazebrook

(*September* 1977)

School of Mathematics
The University of Newcastle Upon Tyne
Newcastle Upon Tyne NE1 7RU
UK

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 195–196
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0195/$02.00/0

Vol. 6
1978

Evaluation of Index Formulae

INTRODUCTION

THE PRIMARY function of an index number is to measure changes in some complex aggregate either through time or from place to place. Familiar examples include index numbers for prices, quantities, and various business activities. A common practice in the construction of such indexes, and the one considered herein, is to weigh the constituent items by quantities which in some sense reflect their relative importance, and then utilize a measure of central tendency to summarize the resulting values. Theoretically, any measure of location can be used as the average of the weighted aggregative items, but because of its apparent economic meaning and its computational simplicity, the arithmetic mean is generally used for this purpose.

A familiar example of the above procedure is the weighted aggregative price index which finds extensive applications in cost-of-living inquiries. In the absence of such unpleasantries as weight revisions and problems of sampling, this index can be simply defined as [1]:

$$I_{i:0} = \frac{\sum_{j=1}^{k} p_{ij}q_j}{\sum_{j=1}^{k} p_{oj}q_j} \tag{1}$$

or by its algebraic equivalent, the value-weighted average of price relatives:

$$I_{i:0} = \frac{\sum_{j=1}^{k} \{p_{oj}q_j\}\frac{p_{ij}}{p_{oj}}}{\sum_{j=1}^{k} p_{oj}q_j}. \tag{2}$$

Here, for each of the k items comprising the aggregate, q_j represents the 'typical' number of units produced, consumed, or the like; p_{oj} is the unit price for the base period; and p_{ij} is the unit price for the specified period i. For convenience, the $p_{ij}q_j$'s will be referred to as the weighted aggregative items. Perusal of the first expression illustrates the characteristics previously noted. In particular, it can be seen (division of numerator and denominator by Σq_j) that the index is a simple ratio of weighted arithmetic means. Or alternatively, if the $p_{ij}q_j$'s are treated as the variables, the index can be interpreted as a ratio of linear combinations of the weighted aggregative variables at two points in time.

The ability of this type of index to summarize the movement of the weighted aggregative items is highly dependent on the distribution pattern of their values [2, 5]. According to studies made on this subject, the use of an index as an average to describe their mass behavior is valid only to the extent that the weighted aggregative items move together through time, i.e. exhibit central tendency. One would therefore surmise that the total variation of these items would represent a major consideration insofar as criteria for the construction and evaluation of index formulae are concerned. However, this is not the case. Proposed properties for index formulae have instead centered on other logical rationales that, perhaps because an index is a binary comparison, basically consider the aggregate at only two points in time. Examples of this class of criterion are the time reversal test and the factor reversal test developed by Fisher [2]. While the properties associated with these tests are desirable, it would appear that the degree of central tendency displayed by the aggregate is a more important consideration, especially in terms of an initial evaluation of the adequacy of this type of index formulae.

THE VARIANCE CRITERION

The construction of an index number requires that a complex numerical aggregate be summarized in terms of a single variable, forcing a k-dimensional system, so to speak, into one dimension. From this viewpoint, a measure of the extent to which the index is capable of representing the movement of the weighted aggregative items through time is the portion of the total variation of the system that it 'explains' [3]. Consequently, if the index is to be defined in terms of a linear combination of the weighted aggregative items, the maximum variation that can be accounted for by a linear function is of particular interest. Knowledge of this quantity would not only provide insight into the ability of this class of index formulae to describe the aggregate, but once determined, it would serve as a basis for evaluating the extent to which members of this class fall short of optimality.

No new theory is required for the determination of the optimum linear function. This can be accomplished via a principal component analysis of the variance–covariance structure of the aggregate. A brief exposition pertaining to the concepts and properties of this statistical procedure will serve to demonstrate its applicability.

Suppose that n observations exist on each of the k variables comprising the vector $\mathbf{x}$ and let Σ denote the covariance matrix for $\mathbf{x}$. The objective of principal components is to derive a new set of k variables, termed principal components, that are linear combinations of the variables of $\mathbf{x}$ subject to the condition that the new variables are uncorrelated [4]. The k-square orthogonal matrix $\mathbf{A}$ required for the transformation $\mathbf{y} = \mathbf{Ax}$ can be expressed as $A = [\mathbf{a}_1\mathbf{a}_2 \ldots \mathbf{a}_k]$ where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ are the eigenvectors corresponding to the latent roots or eigenvalues of the characteristic function of Σ. In addition to being uncorrelated, the principal components possess a number of other interesting and useful properties. In particular, if $\mathbf{a}_i$ is the eigenvector for the latent root λ_i, the variance of the corresponding principal component, $\mathbf{a}_i\mathbf{x}$, is equal to λ_i. Thus, if the principal components are arranged according to the magnitude of their eigenvalues (largest to smallest), the first principal component, $\mathbf{a}_1'\mathbf{x}$ has the largest variance of any linear function of $\mathbf{x}$; $\mathbf{a}_2'\mathbf{x}$ has the largest variance of any linear compound that is uncorrelated with the first compound; $\mathbf{a}_3'\mathbf{x}$ has the largest variance subject to being uncorrelated with the first and second components; and so down to $\mathbf{a}_k'\mathbf{x}$, which is the component with the smallest variance.

Consideration of the above properties reveals that the first principal component provides a solution to

the problem at hand. That is, the largest eigenvalue is the maximum variation of the aggregate that can be explained by a linear combination of the weighted aggregative items. Since the total variation of the system may be expressed as the sum of the eigenvalues, i.e.

$$\sum_{i=1}^{k} \lambda_i,$$

the maximum percentage of the total variation accounted for is simply

$$\lambda_1 \Big/ \sum_{i=1}^{k} \lambda_i, \text{ where } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k.$$

The linear function that explains the maximum variation is of course the first principal component, and it has as its coefficients for the weighted aggregative items the elements of the eigenvector corresponding to λ_1.

AN HYPOTHETICAL EXAMPLE

To illustrate the proposed variance criterion, an hypothetical weighted aggregative price index is considered for the residential usage of certain fuels in the Standard Metropolitan Statistical Area of Baltimore, [1]. The three fuels comprising the aggregate and their respective quantity weights are: Residential gas for use other than heating, 10 therms; electricity, 100 kW h; and residential heating gas, 100 therms. Although these quantity weights are hypothetical and only represent rough estimates of actual consumptions, the price data utilized in the example are those reported in [6]. Prices on each of the three items were observed at 28 points in time: March, June, September, and December for each year in the 7 year period 1970–1976.

Consider now the main question of interest, i.e. what is the maximum variation that can be explained by a linear function of the weighted aggregative items? For the data described above, the covariance matrix for the vector of weighted aggregative items $\mathbf{x}' = [p_{i1}q_1\ p_{i2}q_2\ p_{i3}q_3]$ is

$$\Sigma_x = \begin{bmatrix} 0.2377 & 3.5036 & 1.8922 \\ & 65.5661 & 28.1305 \\ \text{SYM} & & 15.1289 \end{bmatrix}$$

and its eigenvalues are $\lambda_1 = 78.329$, $\lambda_2 = 2.6026$, and $\lambda_3 = 0.0010$. Thus, the total variation of the aggregate and the maximum variation that can be explained by a linear function are $\Sigma\lambda_i = 80.9327$ and $\lambda_1 = 78.329$, respectively. For interpretation purposes, it is perhaps more instructive to note that the optimum linear function is $y = \mathbf{a}'\mathbf{x}$ where $\mathbf{a} = [0.0508\ 0.9118\ 0.4074]$ is the eigenvector for λ_1, account for 96.78% of the total variation in the aggregate. The magnitude of this quantity indicates that the weighted aggregative items move together through time; i.e. they exhibit the strong central tendency that is a prerequisite for the aggregate to be summarized by an index based on a linear function.

Given this result, the maximum explainable variation can be utilized as a reference point for measuring the extent to which the linear function that defines the weighted aggregative price index falls short of optimality. Expressing

$$\sum_{j=1}^{3} p_{ij}q_j$$

as $z = \mathbf{b}'\mathbf{x}$, where $\mathbf{b}' = [1\ 1\ 1]$, the variation the index explains is given by the expression

$$\text{Var}(z) = \frac{1}{\sqrt{\mathbf{b}'\mathbf{b}}}\,\mathbf{b}'\Sigma_x\mathbf{b}.$$

For the example data, $\text{Var}(z) = 49.3285$. Since the total system variation and the maximum explainable variation are 80.9327 and 78.329, respectively, the simple sum of the weighted aggregative items, z, accounts for 60.95% of the total variation and 62.98% of the variation that can be explained by a linear combination of aggregative variables. The interpretation of these results is somewhat subjective. For this example, it appears that the linear function upon which the weighted aggregative price index is based provides only an adequate summary of the movement of the aggregate.

CONCLUSIONS

This paper has considered the evaluation of an index formula from the standpoint of the structure of the aggregate it is to summarize. More specifically, the argument has been made that if the index is to be defined in terms of a linear combination of the weighted aggregative variables, the maximum variation that can be accounted for by a linear function and the variation explained by the linear function defining the index of interest are important measures of the usefulness of the proposed index. For this class of index formulae, it has been shown that the former quantity corresponds to the variance of the first principal component derived from the aggregate's covariance matrix. Once determined, the maximum explainable variation provides not only a basis for evaluating the usefulness of a proposed index, but more importantly, it represents a measure of the ability of this type of index to summarize the aggregate.

REFERENCES

1. *BLS Handbook of Methods* (1971) BLS Bulletin 1711. For sale by the Supt. of Docs., US Govt. Printing Off., Washington, DC (Stock No. 2901–0659).
2. FISHER I (1923) *The Making of Index Numbers.* Houghton Mifflin, New York.
3. KENDALL MG (1974) *A Course in Multivariate Analysis* (lecture notes for a seminar sponsored by the Institute of Advanced Technology). Washington, DC.
4. MORRISON DF (1967) *Multivariate Statistical Methods. Hill, New York.*
5. MUDGETT BD (1951) *Index Numbers.* John Wiley, New York.
6. *Retail Prices and Indexes of Fuels and Utilities.* US Dept of Labor, Bureau of Labor Statistics, Washington, DC.

Ernest B Uhr

Ernest C Houck

(*September* 1977)

Department of Business Administration
College of Business
Virginia Polytechnic Institute and State University
Blacksburg
Virginia 24061
USA

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 2, pp. 197–200
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0501-0197/$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

Bergin JP, Jr (1971) The Relationship of Achievement Motivation to Individual Value Systems. SM Thesis, MIT, Cambridge, MA.

Binkley CH (with **Zitrides TG**) (1971) An Exploration Study of Small Investor Attitudes Toward Investment Media. SM Thesis, MIT, Cambridge, MA.

Bloch JG (1971) Feasibility Study of Offshore Facilities in Massachusetts Bay for Unloading Residual Fuel Oil. SM Thesis, MIT, Cambridge, MA.

Blount FW (1971) Public Utility Regulation and its Effect on the Managerial Style and Innovative Incentive of the Regulated Firm. SM Thesis, MIT, Cambridge, MA.

Boisot MH (1971) The Architectural Profession: A Management Viewpoint. SM Thesis, MIT, Cambridge, MA.

Botteri GM (1971) The Impact of the External Environment (Including New Technology) on the Internal Environment of the Massachusetts Institute of Technology. SM Thesis, MIT, Cambridge, MA.

Buchanan CA, JR (1973) Development Banking Model—A Framework for Analysis. SM Thesis, MIT, Cambridge, MA.

Clemot MG (1973) Offshore Petroleum and Gas Supply Technological and Economic Trends. SM Thesis, MIT, Cambridge, MA.

Cumberford J (1973) The Underlying Factors of the Structural Changes in the Venture Capital Markets. SM Thesis, MIT, Cambridge, MA.

Diederich JL (with **Packard PK**) (1971) Venture Management in the Industrial Sector. SM Thesis, MIT, Cambridge, MA.

Dieguez GA (1971) The Bank Acquisition Process. SM Thesis, MIT, Cambridge, MA.

Dietrich WK (1971) Factors Affecting the Forward Integration Policies of the Relative Newcomers on the International Oil Scene. SM Thesis, MIT, Cambridge, MA.

Dix MW, Jr (1973) The Assistant District Attorney: A Participant Observation Perspective. SM Thesis, MIT, Cambridge, MA.

Dooley TM (1971) An Analysis of Factors Associated with Innovation and Community Support in Massachusetts Public School Systems. SM Thesis, MIT, Cambridge, MA.

Egashira K (1971) Marketing Issues in a Social System: Prevention of Venereal Disease. SM Thesis, MIT, Cambridge, MA.

Faucett RB (1971) The Management of Venture Capital Investment Companies. SM Thesis, MIT, Cambridge, MA.

Fifer RB (1971) Cost of Capital: An Analysis of the Practical Applications for the Firm. SM Thesis, MIT, Cambridge, MA.

Flasaquier M (1971) Planning and Control in Non-Profit Organizations. SM Thesis, MIT, Cambridge, MA.

Frankel RM (1971) A Normative Approach to Strategy Formulation for Survival in the Brokerage Industry. SM Thesis, MIT, Cambridge, MA.

French JL, Jr (1971) Teachers' Attitudes Toward Teacher Aides. SM Thesis, MIT, Cambridge, MA.

Freytag RA (1971) Foreign Financial Investment by United States Banks: The Role of the Bank Holding Company. SM Thesis, MIT, Cambridge, MA.

Frondistou-Yannas SA (1973) Dynamics of the Precast Concrete Industry. SM Thesis, MIT, Cambridge, MA.

Fry RE (with **Lech B**) (1971) Organizational Development Approach to Improving the Effectiveness of Neighborhood Health Care Teams: A Pilot Program. SM Thesis, MIT, Cambridge, MA.

Fuchs GA (1971) IATA and its Regulators. SM Thesis, MIT, Cambridge, MA.

Gallardo-Kuri AJ (1971) Correlates of Giving to the M.I.T. Alumni Fund. SM Thesis, MIT, Cambridge, MA.

Gehman JR, Jr (1971) A Study of Computer Program Acquisition Management in Several Air Force Projects. SM Thesis, MIT, Cambridge, MA.

Gnecco-Lombardi ML (1971) Congruence of Job Factors in Brazilian Development Banks. SM Thesis, MIT, Cambridge, MA.

HAN VT (1973) The Export Processing Zone and South Vietnam After the War. SM Thesis, MIT, Cambridge, MA.

HARRIS GI (1973) An Analysis of Technical Communications Within a Group of Five Research and Development Laboratories. SM Thesis, MIT, Cambridge, MA.

HEDGE AJ (1973) The Strategic Process in a Academic Medical Center. SM Thesis, MIT, Cambridge, MA.

HERBERT WR III (1973) An Examination of Significant Internal Growth in the Aerospace Corporate Environment. SM Thesis, MIT, Cambridge, MA.

HOLLEY MJ III (1973) Information Handling in a Local Government Public Department. SM Thesis, MIT, Cambridge, MA.

HOLT GA JR (1971) Semantic Models for Data Description. SM Thesis, MIT, Cambridge, MA.

HOOPER JO (1973) Foreign Investment Decision Techniques Practiced by Multinational Companies Specifically in Planning, Personnel and Financial Management. SM Thesis, MIT, Cambridge, MA.

HORTON RB (1971) The Utility of Models in the Planning Process. SM Thesis, MIT, Cambridge, MA.

INADOMI RJ (1971) Attitudes toward Change: A Case Study of Computer Systems Simulation. SM Thesis, MIT, Cambridge, MA.

KAISER JG (1973) The Decision and Evaluation Processes of a Proposed Merger or Acquisition. SM Thesis, MIT, Cambridge, MA.

KNOWLES RC, JR (1973) Performance Evaluation Feedback in the United States Marine Corps—A Critical Analysis. SM Thesis, MIT, Cambridge, MA.

KUBOTA T (1973) Practical Management Systems in Japan. SM Thesis, MIT, Cambridge, MA.

KUPASRIMONKOL S (1971) Short-Term Financial Planning for a Multinational Enterprise under Unstable Monetary Conditions—A Systematic Analysis. SM Thesis, MIT, Cambridge, MA.

LACROIX RLJ (1971) Opportunity Surveys: A Model for Comparative Judgement. SM Thesis, MIT, Cambridge, MA.

LAFFARGUE JM (1971) Intracompany Pricing Policy of a European Company Having a Subsidiary in North America. SM Thesis, MIT, Cambridge, MA.

LAVIGNE JC (with PORGES D) (1971) A Management Information and Control System for a School and its Applications to the Sloan School of Management. SM Thesis, MIT, Cambridge, MA.

LEBNER M (1971) Risk Averseness, Brand Sensitivity Innovativeness, and a Model of Buyer Behavior for New Products. SM Thesis, MIT, Cambridge, MA.

LENTZ C (1971) Project Planning Model for a Consulting Engineering Firm. SM Thesis, MIT, Cambridge, MA.

LEONARD MA (1971) Factors and Requirements for the Improvement of Automotive Manufacturing Quality. SM Thesis, MIT, Cambridge, MA.

LEWIS BR (1971) Do the Poor Pay More? SM Thesis, MIT, Cambridge, MA.

LISTER RG (1973) A Systems Framework for Marketing Communications Equipment: A Case Study of Audio Response. SM Thesis, MIT, Cambridge, MA.

LIVELY MB (1971) Time Aspects of Paging on Multics. SM Thesis, MIT, Cambridge, MA.

LLOYD K (1971) Management of the Data Processing Department. SM Thesis, MIT, Cambridge, MA.

MACCOWATT HP (1973) Financial Disclosure: Practices and Problems. SM Thesis, MIT, Cambridge, MA.

MALBRANCKE J (1971) Implications of the Decision Approach to the Software Industry. SM Thesis, MIT, Cambridge, MA.

MANCHOLA-ARANA G (1973) A Playground for Foreign Investments? SM Thesis, MIT, Cambridge, MA.

MASSON P (with MOODY TC and STUBBS JD) (1971) Planning and Control for Community Hospitals: A Case Study of the Cambridge Hospital. SM Thesis, MIT, Cambridge, MA.

MATHEWS ER (1971) A Proposed Management Development Program for a Government R & D Center. SM Thesis, MIT, Cambridge, MA.

MCCARTER WT (1971) The Utilization and Development of the Technical Communications System in a Research Organization. SM Thesis, MIT, Cambridge, MA.

MONRO-DAVIES ARD (1971) An Investigation into Maine Potato Future Prices. SM Thesis, MIT, Cambridge, MA.

MURRAY MH (1971) A Study of Computer-Output-Microfilm Systems. SM Thesis, MIT, Cambridge, MA.

MYERS (1971) Management Practices in Non-Union Companies. SM Thesis, MIT, Cambridge, MA.

NORA HJ (1971) An Automated Medical History Questionnaire: A Study of its Use and Acceptance by Physicians. SM Thesis, MIT, Cambridge, MA.

ORAM PB (1971) Public Law 87-653: Truth in Negotiation—An Evaluation. SM Thesis, MIT, Cambridge, MA.

PACE SC (1971) A Concept of Strategic Industry Analysis with an Application to the Snowmobile Industry. SM Thesis, MIT, Cambridge, MA.

PETERS JH (1971) A Health Dynamics Model—Phase I. SM Thesis, MIT, Cambridge, MA.

Polishuk P (1971) Transferability of Aerospace Management Skills to the Industrialized Housing Industry. SM Thesis, MIT, Cambridge, MA.

Poor RM (1917) White Business after the Riots. SM Thesis, MIT, Cambridge, MA.

Popper J (1971) Buffer Stocks for Stabilization of Commodity Markets: A Dynamic Study. SM Thesis, MIT, Cambridge, MA.

Preiss GW (1971) Work Goals of Engineers: A Comparative Study Between German and US Industry. SM Thesis, MIT, Cambridge, MA.

Rothman JS (1971) An Analysis of Automated Medical Record Systems. SM Thesis, MIT, Cambridge, MA.

Smith SC (1971) Planning for People in MIS Development. SM Thesis, MIT, Cambridge, MA.

Smith ER, Jr (1971) Student Decision Making Models in the Graduate Business Environment. SM Thesis, MIT, Cambridge, MA.

Smith LM (1971) A Vertical Profile of Management Attitudes Toward People, Policies, and Practice. SM Thesis, MIT, Cambridge, MA.

Sperling AL (1971) Unit Dose Drug Distribution: A Revolutionary Approach to Better Health Care. SM Thesis, MIT, Cambridge, MA.

Spooner TB (1971) An Analysis of Prepaid Group Practice. SM Thesis, MIT, Cambridge, MA.

Stergion PB (1971) Territory Control System. SM Thesis, MIT, Cambridge, MA.

Steuert JJ (1971) Design of a Mathematical Stochastic Simulation Capability. SM Thesis, MIT, Cambridge, MA.

Swain RL (1971) Catalytic Colleagues in a Government R & D Organization. SM Thesis, MIT, Cambridge, MA.

Taie AG (1971) New Business Development: A Case Study of the Mini Computer Industry. SM Thesis, MIT, Cambridge, MA.

Thiel ML (1971) Ireland's Industrial Development Program: A Survey of its Impact on US Corporations. SM Thesis, MIT, Cambridge, MA.

Thomas JA (1971) Implementation of an Automated System into the Bank Lending Function with Respect to Small and Medium Sized Firms. SM Thesis, MIT, Cambridge, MA.

Torpie AL (1971) Capital Budgeting and Appropriation as a Planning and Management Control System. SM Thesis, MIT, Cambridge, MA.

Traub AC, Jr (1971) Design of a Commercial Loan Information System. SM Thesis, MIT, Cambridge, MA.

Turner PW (1971) Information Systems: A Framework for Analysis. SM Thesis, MIT, Cambridge, MA.

Vallhollebeke JGM (1971) Establishment of a Single European Currency in the European Economic Community. SM Thesis, MIT, Cambridge, MA.

Walker ET (1971) A Program Budget for a Neighborhood Health Center. SM Thesis, MIT, Cambridge, MA.

Wall KE, Jr (1971) Reducing 'No-Shows' in a Large Group Practice Medical Outpatient Facility. SM Thesis, MIT, Cambridge, MA.

Wallitsch FW (1971) A Model for the Measurement of the Effectiveness of Use of a Plant's Physical Resources in Manufacturing. SM Thesis, MIT, Cambridge, MA.

Weber J (1971) Development of the Health Record and Information System for the Boston Model City's Family Life Centers. SM Thesis, MIT, Cambridge, MA.

Weissinger RH (1971) Motivational Determinants of Production Rates: A Test of the Galbraith–Vroom Model. SM Thesis, MIT, Cambridge, MA.

Wetmore WC (1971) Models of the Firm's Transactions Demand for Cash: A Survey and Extensions. SM Thesis, MIT, Cambridge, MA.

Whetstone JT, III (1971) The Quantitative Investigation of the Feasibility of and Best Location for Regional Check Clearing Centers. SM Thesis, MIT, Cambridge, MA.

White AF (1971) Preparation of Managers for Cross-Cultural Assignments. SM Thesis, MIT, Cambridge, MA.

White JD (1971) Cross-Perceptions of Research Engineers and Administrators. SM Thesis, MIT, Cambridge, MA.

Whitmore JC (1971) Resource Utilization in Time-Sharing Computer Systems. SM Thesis, MIT, Cambridge, MA.

Wiesner FA (1971) Learning Profiles and Managerial Styles of Managers. SM Thesis, MIT, Cambridge, MA.

Wilk RA (1971) Informational Content of Changes in Corporate Dividends and Earnings. SM Thesis, MIT, Cambridge, MA.

Wilkes LL, III (1971) A Model Based Analysis of Inventory Shrinkage Control in Retail Stores. SM Thesis, MIT, Cambridge, MA.

Williams SJ (1971) The Operational Aspects of a Medical System for Abortion Services. SM Thesis, MIT, Cambridge, MA.

Wilson KW (1971) The Aviation Safety System. SM Thesis, MIT, Cambridge, MA.

Vol. 6
1978

WOLF FL (1971) Cooperatives and Condominiums: A Financial Analysis. SM Thesis, MIT, Cambridge, MA.

YORK BW (1971) A SIMPL Simulation of ASP Scheduling. SM Thesis, MIT, Cambridge, MA.

YUCKMAN DA (1971) Collective Bargaining for University Facilities. SM Thesis, MIT, Cambridge, MA.

The following are obtainable at the Lockwood Library, State University of New York at Buffalo, 3435 Main Street, Buffalo, New York 14214, USA:

ALLEN R (1976) The Labor Market Experiences of Two-Year College Alumni. PhD Thesis, University of New York at Buffalo.

CADY J (1975) Structural and Competitive Effects of Retail Trade Regulation: A Study of the Impact of Public Policy in the Retail Market for Prescription Drugs. PhD Thesis, University of New York at Buffalo.

CUMMING P (1976) National Blood Policy and the American Blood Commission: A Systems Study. PhD Thesis, University of New York at Buffalo.

GOMOLKA E (1974) The Effect of Personality Systems of Small Business Company Presidents as a Determinant of Organization Structure. PhD Thesis, University of New York at Buffalo.

HIERL A (1974) The Effects of Warrant Financing on the Valuation of the Firm. PhD Thesis, University of New York at Buffalo.

KENDALL K (1975) A Systems Study of Blood Inventory Management. PhD Thesis, University of New York at Buffalo.

KHAN M (1975) Police Service System. PhD Thesis, University of New York at Buffalo.

KIM EH (1975) A Theory of Optimal Financial Structure in Market Equilibrium: A Critical Examination of the Effects of Bankrupty & Corporate Income Taxation. PhD Thesis, University of New York at Buffalo.

POST J (1976) Risk and Response: Management and Social Change in the American Insurance Industry. PhD Thesis, University of New York at Buffalo.

RULAND W (1976) The Accuracy and Information Content of Earnings per Share Forecast Reports by Firms. PhD Thesis, University of New York at Buffalo.

SENBET L (1975) International Capital Market Equilibrium: Theoretical Analysis and Empirical Investigation. PhD Thesis, University of New York at Buffalo.

STEVENS J (1976) Managerial Commitment, Policy Receptivity, and Policy Implementation in Public Sector Organizations. PhD Thesis, University of New York at Buffalo.

WELKE R (1975) A Conceptual Framework and Methodology for Describing and Evaluating Alternative Information Systems from a User or Organizational Perspective. PhD Thesis, University of New York at Buffalo.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 201–205
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0201$02.00/0

Vol. 6
1978

EDITORIAL

Political Man

BUDDING young managers, flushed with enthusiasm as they emerge from sophisticated business schools, tend to regard themselves as the undisputed custodians of the management scene. Their ethos is that management is a professional activity, where skill and knowhow are essential ingredients in the conduct of dispassionate analyses of problems, so that both symptoms and causes can be identified as a basis for the decision making process. They have been taught that management is concerned with the use of resources and that analytical tools are available for carrying out a proper evaluation of alternative courses of action. It therefore follows that, given a reliable data bank and stated objectives, new recruits to the managerial ranks should be adequately qualified to construct decision models, to feed the relevant information into them, and to draw the appropriate conclusions.

This may be regarded by many as a somewhat exaggerated mechanistic outlook, and in certain respects it is, though the term 'mechanistic' should not necessarily have a derogatory connotation in this context. The vision of the young manager is that of becoming a master of the tools of his particular trade, first by understanding the theoretical foundations of these tools, secondly by realizing their general applicability, and thirdly by learning how to use them in real situations. It is a healthy vision, shared by other professionals and technicians alike, and the fact that many managers are immersed in tasks which are well-defined, and sometimes totally prescribed, should in no way detract from their importance. A great number of managers (and analysts) take pride in being good technicians, and rightly so: much of the routine work of most enterprises depends on the technical skill and on the ability of these managers to keep the organization going.[1]

But, clearly, training managers to become good technicians is not enough. The prescribed routine tasks predominate at the lower echelons of a managerial hierarchy, but as managers ascend the pyramid, they increasingly see the necessity to re-examine the nature of the prescriptions imposed on the managerial system and reflect on their wider consequences. It is not so much that they doubt the validity of the decision models in use, but they begin to question what goes into them: the assumptions, the relationships between variables, the organizational constraints, the reliability and validity of the data, and above all—the definition of objectives and performance criteria. It is, therefore, natural for these issues to be exposed and debated at some length in business schools, since it is generally assumed that business graduates are likely to progress during their careers to positions of responsibility, where questions of this kind inevitably become pertinent.

It is interesting to note this trend in OR (Operational Research) courses as well. The large majority of these courses have been—and alas some still are—utterly obsessed with mathematical formulations and algorithmic pur-

[1] See also the editorial "Technician or Adviser", *Omega* **5** (1977), 1–6.

suits, following from the assertion that most managerial problems are amenable to quantitative analysis. Thus, the notion has long persisted that if you can fashion a problem in rigorous mathematical terms, together with its attendant solution procedure, you have done a good job as an analyst, or as a manager, and that this is what the major part of the management task is all about. Time and experience have taught us, however, that (at best) models are weak simplifications of reality, that a proper appreciation of problems cannot be detached from the environment in which they occur, and that the knowledge of sophisticated techniques is a necessary but not sufficient condition for OR to be successful. Hence we find that many OR courses now include foundation subjects, coupled with a concern for measurement, implementation and the practice (as opposed to the theory) of OR. Scrutiny of the value of the information put into models, coupled with questions on objectives, now form an integral part of an increasing number of OR courses. Thus, the realization that the total management task requires more than the bare tools of the technician is not confined to the business schools. We may think that this should have been obvious all along, but the process of evolution—and it is far from being universal or complete—has been painful nonetheless, with OR centres showing perhaps a greater reluctance to adapt than the business schools.

What is quite clear from changes in the business environment over the last decade or two is that this process of adaptation, significant as it has undoubtedly been, has not gone far enough. As many senior managers will readily testify, the managerial scene is characterized nowadays by a multitude of glaring and often baffling contradictions. This point was forcefully brought out in a recent report entitled "Educational and Training Needs of European Managers", published last year by EFMD, namely the European Foundation for Management Development (though there is little in the report which is intrinsically European and which is not applicable to most developed countries in the Western World).

Vol. 6
1978

First, the report argues that the very success of the business enterprise has led to further demands upon it from society and to greater expectations that the firm should be able to meet all the aspirations of its members. The success hitherto of the enterprise as an economic system is characterized in the report "as the motor of the prosperity of society, as an agent for the mobility and the development of individual talent, and as a stimulant of the 'open society'", but the ensuing changes in social values and expectations have led to "hostility of many groups in society towards corporations (which) can be partly explained by the unwillingness of corporations to look at the non-economic repercussions of their consideration of man as *homo oeconomicus*", and "social reformers demand a total systems approach to the change of working environments as a necessary prerequisite to desirable changes in society".

The world of the manager is now beset by the challenge of contradictions. These contradictions are not just a reflection of internal conflicts between performance criteria, such as the characteristic battles between the sales, finance and production functions, which require the analyst to handle models with multiple-objectives. The contradictions are at a much more fundamental level, in that they challenge the very *raison d'être* of the enterprise and its place in society. In particular, the contradictions centre on the following issues:

- economic criteria *vs* social responsibility
- technology *vs* employment

- national *vs* multi-national interests
- ownership and participation *vs* unitary control
- reactive role *vs* proactive role.

The conflict between economic criteria and social responsibility is at the heart of the matter. It is commonplace nowadays to assert that the dictum of *economic man* should give way to that of *social man*, but while it is generally accepted that a business enterprise has obligations to people, the full extent and implications of adopting a social responsibility posture are far from obvious. Economic man will see his prime role as that of advocating courses of action solely on economic criteria, with various social considerations regarded as constraints, although even some of these constraints are said to be rooted in economic rationality: you provide better rewards and facilities for your employees, not just because of social pressures or fear of losing good people, but because you believe that the extra expenditure will lead to improved performance, either in the short or in the long term. Thus, economic man will often rationalize his actions on economic criteria, though he may have to concede that there are occasions when actions prompted by social or political considerations have no financial justification.

In contrast, the stance of social man is that such actions are a matter of conscience: the enterprise should undertake them not because it is under coercion to do so, but because it is right. Social responsibility stems from the premise that the enterprise exists in order to fulfil the needs of people, not only of its employees or customers, but also of the inhabitants of the locality in which it is situated, and of society at large. It is therefore incumbent on the management of an enterprise to seek ways of improving the environment and the general quality of life. There are, of course, various constraints which social man has to bear in mind, the primary one being that the enterprise must survive (if it is to continue to bestow benefits on society) and hence must heed the rigours of the market-place and its harsh realities. What is not clear, however, is how much social man is prepared to concede on economic issues (and all social men are not, of course unanimous on this question). For example, he may accept that it is essential for a business enterprise to make a profit, but views do differ on the minimum level of profit that must be attained and on what use this profit should be put to.[2]

Perhaps the most prominent area in which the theories of economic man and social man clash concerns the use of technology. If it can be shown that new technology will make a plant more efficient in the use of resources, will cut down unit costs, will improve productivity, but require a smaller work-force, should it be embraced? Three alternatives need to be considered:

- adopt the new technology and reduce the workforce
- adopt the new technology but keep the workforce intact
- stick to the old technology.

Economic man will advocate the first alternative, arguing that the other two are bound to result in loss of competitiveness and hence will threaten the long term survival of the plant. It has even been suggested that higher

[2] See also the editorial "On the Corporate Ethos", *Omega* **1** (1973), 393–396.

technology is a net generator of employment (the work of J Kendrick, Chief Economist of the US Department of Commerce is cited in the EFMD Report in support of this assertion), so that while a given plant may require less manpower for a given level of output, it may be able to increase its capacity; also, plants in new industries will emerge and the services sector will expand to more than compensate for the immediate loss of jobs. The counter argument is to question the full effectiveness of this compensatory mechanism (let alone the unavoidable time lag that is involved), to cite the relatively high levels of unemployment and to point to the very large capital expenditure that is required nowadays in advanced technology to generate even a modest number of new jobs. This is why social man tends to prefer the third alternative and to suggest that the unbridled adoption of new technology can have dire social and economic consequences. The second alternative, which accepts the new technology but tolerates over-manning, is regarded by some as a reasonable compromise, in that it allows the immediate effects to be cushioned until new job opportunities arise, and it also raises the possibility of a reduction in the length of the working week, though the problem of reduced competitiveness remains, in comparison with plants that adopt either the first or the third alternative.

Another example that illustrates the clash between the two philosophies is the position of a multi-national company. Whatever decisions it makes in the allocation of its resources, it is bound to benefit one country and thereby discomfort another: expanding production or opening a plant in country *X* rather than in country *Y*, even producing in *X* with the intention of selling in *Y*, are seen by the people of *Y* as a denial of employment opportunities. Furthermore, the movement of profits and dividends from one country to another, with a view to achieving a reduction in the total tax liability (and which multi-national corporation is not engaged in such exercises?) results in certain governments and their nationals being aggrieved by the deprivation of tax income. Should a corporation operating in a given country feel morally obliged to pay all the taxes it can in that country, as a contribution to the total social good, or should it avoid payment by all feasible legal means with the primary objective of looking after the interests of its shareholders and employees?

The questions of ownership, participation in profits, and the growing clamour for participation in decision making (under the guise of industrial democracy) have recently been the subject of vociferous debates. The protagonists range from those who regard the declining role of the shareholders as not only inevitable but desirable, to those who uphold the need for unitary control and who see its dilution as a serious threat to the ability of management to take decisions quickly and efficiently. Whichever side you are on, you have to acknowledge that changes in the climate of opinion have taken place, that contradictions and conflicts do exist, and that multi-objective modelling for resolving them are hardly likely to be very effective.

It is quite evident that over the last two decades the management task has become increasingly political in character, in terms of grappling with ideologies and of the practical necessity to get things done in the face of mounting difficulties. The modern manager needs not only to have the ability to analyse a problem and arrive at a solution, not only to persuade and communicate, but also to negotiate, to handle pressure groups, to assess the internal and external political climates, to know when to compromise and when to stand firm. Economic man and social man have given way to *political man.*

Vol. 6
1978

One of the basic tasks of political man is to decide when the enterprise should be *reactive* and when it should be *proactive*. In a reactive role, it merely responds to pressures, whether they are external (strong public opinion, or the outcome of legislation) or whether they emanate internally from the workforce. This passive approach is compatible with the view that by and large there is no obligation for business to undertake commitments for the betterment of society or the environment which should remain the responsibilities of local or central government. In a proactive role the enterprise takes on a mission to induce changes in society, and in exercising its initiative it tries not only to anticipate the changes that will occur in any event, but also to influence the pattern of development and adaptation.

The reactive role is one which managers readily understand. They constantly have to face changes in market conditions, in the actions of their competitors, in technology and innovation, and in demands from trade unions; whenever new conditions arise, they have to assess the situation and determine the most appropriate response. It is only natural for this attitude to be extended to the handling of social or environmental demands: do not anticipate them, act only when you have to. Now, managers also have experience when it comes to a proactive role, such as when decisions are made to open new plants, to launch new products, to use new methods of production or to enter new markets. But these initiatives are confined to business and technical affairs. The proposition that it can, or should be, extended to initiating changes in the social environment has met with criticism from many sceptics, who question whether a business enterprise has the moral right to presume that it knows what is good for society, and furthermore that it has the obligation to employ resources to that end. There may also be vociferous resentment to what might be considered a return to a paternalistic stance.

By definition, there is no-one who can offer a complete body of knowledge and a set of prescriptive answers to these many questions. What is clear, however, is that the problems exist, that their political connotations cannot be ignored, and that they increasingly impinge on the manager and on his task. EFMD were right, therefore, to highlight some of these issues and to refer to their implications on curricula in business schools. But perhaps their most important conclusion is that: "Corporations should see that managers are given ample opportunities for self-development. This calls for experience in different jobs and at different levels, training in line with an individual's changing needs, and education, which has a longer-term objective. Throughout his working life there should be developed, in discussion and by agreement with him, a phased programme of training and development opportunities extending from recruitment well into his 50's, in the form of '*éducation permanente*'".

SAMUEL EILON
Chief Editor

BLANK
PAGE

Vol.
197

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 207–209
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0207$02.00/0

Vol. 6
1978

FEEDBACK

Multidimension Scaling for Multiobjective Policies: A Comment

PROFESSOR Rivett in his recent paper [2] may well have opened up new ground in the development of techniques for deriving and representing preference structures. He has done this by recognising the potential of methods developed by Kendall and others for deriving structure in a variety of areas. However it is possible at this stage to suggest situations where the method would be unsuitable or at least inefficient.

Rivett's method is based on a binary indifference relation (I) defined over a set of policies (P) which is analogous to Kendall's abuttal relation (A) defined over a set of geographical entities e.g. departments of France, states of the USA (D). The important property which these two relations share is non-transivity. This is obviously necessary for (A, D) and is possibly desirable for (I, P) since many convincing arguments have been proposed for so regarding indifference e.g. Fishburn [1]. Whilst it is appropriate to permit indifference to be non-transitive this is not to deny that for some decision maker confronted with some set of policies P, his indifferences will be transitive. It is in situations such as this that Rivett's method is weak.

If (I, P) is transitive, symmetric (which Rivett assumes) and reflexive (which is clearly reasonable) I will generate an exhaustive set of mutually exclusive subsets (equivalence classes) on P. I will hold between all pairs within subsets but not between any pairs in different subsets. In this case the algorithm will not be able to distinguish between alternatives within subsets and will not be able to locate the subsets relative to each other.

Actually whilst transitivity is sufficient to cause problems, nontransitivity is not sufficient to resolve them. It is quite easy to conceive of situations where P is partitioned by I into mutually exclusive subsets which are not 'linked' but I does not necessarily hold between all pairs within a subset. This can be demonstrated using Rivett's robustness problem. For example, rewrite Criterion 1 as

$$i\, I\, j \text{ when } |v_i - v_j| \leqslant k.$$

If $k = 5$ then the set of twenty five alternatives would be partitioned into four subsets: 1–13, 14–19, 20–22 and 23–25. Only the last of these is transitive. The algorithm would be able to produce 'maps' for each of these subsets but not relate the subsets to each other in an overall preference structure.

A primary requirement for Rivett's method to work effectively is therefore that the graph corresponding to (I, P) be connected (3) and non-transitivity is *necessary*, but not sufficient, for this.

In any application, whether (I, P) is connected or not can be determined by examining the reachability matrix (4) of the graph. This can easily be computed from the original indifference data. Thus situations where the method is likely to prove unfruitful can be screened out in advance; although if this test is passed there is no guarantee of effectiveness.

Returning to Rivett's Criterion 1 robustness test it is clear that for values of k greater than or equal to the greatest distance between adjacent policies the graph of (I, P) will be connected and vice versa. The best reproduction of the one-dimensional trial map will be achieved when k is equal to the greatest distance between consecutive alternatives (i.e. $k = 6$). In this case connectedness is assured yet the effect of 'spurious' links minimised.

Taking the reported robustness tests as a whole it would certainly appear that the method is promising. More importantly, perhaps, on the basis of the comparison between this method and that requiring a multiattribute utility function it seems that the latter may well be robust despite how one feels about the assumptions on which it is based. What is used in practice would then depend on personal taste and operational convenience.

REFERENCES

1. FISHBURN PC (1970) Intransitive indifference in preference theory: A survey. *Op. Res.* **18**, 207–228.
2. RIVETT BHP (1977) Multidimension scaling for multiobjective policies. *Omega* **5**, (4), 367–379.
3. ROBERTS FS (1976) *Discrete Mathematical Models*. Prentice–Hall, New York.
4. ROBERTS FS (1976) *Discrete Mathematical Models*, pp. 56–58. Prentice–Hall, New York.

RH GREEN

School of Management
University of Bath
Claverton Down
Bath BA2 7AY
UK

On The Role Of Operational Research

DANDO and Sharp have expressed the hope [4] that Eilon's recent editorial, "Technician or Adviser" [7] will trigger a major debate on the proper scope of OR. Whilst sharing that hope, at the same time I feel it necessary to question not only the scope of OR but the nature of the function as well. This contribution to the debate, like that of Dando and Sharp, probably also strays from the paths normally frequented by the OR community and is similarly intended not to narrow, but to widen that debate.

Since one of the universally accepted features of the OR function is its model building activity and since, not infrequently, this activity is so zealously pursued that the result is not entirely appropriate to the real problem [2], I shall begin by closely examining the definition (or model) or OR adopted by the OR Society. Firstly I suggest that the term 'distinctive approach' used in the definition betrays the real purpose of adopting it. Rather than to clarify the nature of OR the purpose, I suggest, was to establish clear lines of demarcation with other disciplines with which, at the time, the OR community was, or perceived itself to be, in competition. The value system associated with such a purpose is, I believe, not one entirely compatible with the synoptic view attributed by Eilon to an 'adviser' nor with the view of OR as an interdisciplinary activity.

Secondly, perhaps the worry many OR workers have over the adequacy and relevance of much of their work [5] arises because of a difference between the model of OR in the minds of the OR workers and that in the minds of many of their client decision makers, managers or organisations. The latter's model of the OR worker, albeit implicit, is perhaps more often akin to that of an organisational tinker who performs a wide variety of the functions essential for the efficient working of an organisation but for which, individually, it would be uneconomic or impolitic to warrant a full time employee. The most obvious functions which spring to mind from personal experience are facilitating the flow of information across organisational boundaries and the improvement of information systems. However, in the case of other functions, such as defining the real problem, clarifying objectives and arbitration between rival techniques, the role might more properly be described as that of a 'professed fool'.

On the other hand the OR worker's own conception of his activity is largely responsible for his preoccupation with mathematical modelling, which in extreme cases amounts almost to a fetish. This is readily understandable when one remembers that war gaming, which was an early form of OR model, is now one of the fastest growing leisure activities. Quite simply, modelling is infectious fun. For this reason, quite apart from Dando and Sharp's distinction between 'convergers' and 'divergers', it is very difficult to persuade a technician to give up his paid hobby and become an adviser. This difficulty will be appreciated most by those who have observed how the model world becomes a reality and affects the behaviour of role-playing game participants.

The OR society definition acknowledges that the complex systems to which OR is applied include men, but does not appear to view them differently from the machines, materials and money also listed in the definition. However, because of this involvement Dando and Sharp [6] have questioned the myth of OR as a physical science and have instead suggested that a social science is needed to tackle such problems. Because many of the important features of the problem originate in the minds of the men acting as decision makers it is self-evident that the establishment of effective communication between OR worker and client is a necessary prerequisite for successful problem formulation.

But OR is not the only activity involving human problem solving. Shaw [9] sets out a basic counselling model which is strikingly similar to that put forward as a general model of the OR process [1]—this is not surprising if one takes the broader view of OR as an Inquiring System [3]. And the counselling model, like that of OR, is acknowledged as a simple generalisation of what is essentially a dynamic and adaptive process. It is interesting too, that recent research [10] aimed at exploring concepts of counselling found a degree of difficulty in defining this function similar to that experienced by the OR community in defining theirs.

However, Law [8] has emphasised the unifying elements of the various counselling professions he studied. In as much as the OR process shares the general form of model, I personally prefer to view OR as a form of organisational counselling in which explicit (but not necessarily uniquely mathematical) modelling is the special skill practised. Such a view, I believe, acknowledges the special role of the OR worker in interacting with the human decision makers in the client system (while perhaps avoiding the connotations of spontaneous short term assistance and of 'Delphic wisdom' which might be inadvertently implied by the term 'adviser'). It also emphasises the need to develop revised attitudes towards the OR process itself, including quite different measures of its effectiveness, as well as the need for training programmes less grotesquely biased towards mathematical techniques.

The most significant implication however is the organisation's 'ownership' of the problem — this renders the 'backroom scientist' image of OR totally inapplicable. No 'solutions' can be promised to any individual decision maker within the organisation. These can only be arrived at, if at all, by a continuing and wide-ranging dialogue between all the Decision Makers in the organisation and the OR worker in which the skills of the OR worker assist the organisation as a whole in articulating and thus encapsulating the important features of its world. This can present problems for example when the OR group is attached to, or embedded in, a particular functional element of the organisation, or when one particular decision maker regards himself as the prime client and rejects the broader model for parochial reasons.

There are however, two particular points which are very relevant to the debate on the role of OR. Firstly, I have already mentioned the strong attraction of the model world in the context of role-playing. It is equally true that the OR worker can all too easily forget that his models are, and can never be more than, oversimplified models of reality. If there is any danger at all in adopting a counselling style of OR it is that, in attempting to work so closely with the decision makers in the client organisation, the eager OR worker may persuade them to lose sight of the distinction as well. However a common situation occurs when the OR worker can

defend his model by logical argument based on accepted facts and yet the organisation, or particular and important decision makers within it, are still unwilling to adopt the new perspective. What should the OR worker do in these circumstances? Should he discard his model and tackle a new problem? The answer must be to maintain the debate, possibly by searching for other models less radical which might form 'stepping stones' to the eventual adoption of the original model. Here the skills of rationalising, not rejecting, the 'irrationality' of the decision makers is paramount and psychology (and other allied disciplines) are more relevant than mathematical logic.

Secondly, even if the OR worker can be relied upon to present his results impartially and unbiassed by his own values (and we must not underestimate the difficulties of doing this when we are involved in problems which have wide ranging social implications for the whole community) there is an even subtler danger which must be acknowledged. White [11] in developing a model of the OR process, has raised the issue of 'secondary decisions' and it must be remembered that these may have a very significant effect on the behaviour of the model produced. They, too, therefore may properly be the province of the decision-makers in the organisation, albeit with professional guidance from the OR worker. The need, in this case, is to develop plural models so that the debate can also encompass this dimension.

One of the main skills of the counsellor is 'listening'. I believe that if we listen more to all the members of the organisation, pay less attention to the text books filled with standard techniques and resist the pressures from our line managers and peer groups to continue to practice OR as a purely mathematical exercise aimed at local optimisation we will be able to make a more effective contribution to the problems of the organisation than we are managing to do at present. This would at least free us from, *inter alia*, the time consuming activities of large-scale computer model building, and so give us more time to spend with the decision makers in the system. However I do not imply that by listening more we should adopt a passive role. Quite the opposite. By widening his base of experience and understanding the OR worker can speak louder and more confidently. We would aim to offer the decision makers a wider range of more appropriate (albeit sometimes unpalatable) models on which to reflect. To do this it is necessary for the OR worker to gain the confidence and respect of all the decision makers while at the same time being accepted by them as independent and non-partisan. But the hardest job of all will be changing the expectations of OR held by others, and by ourselves. This will be particularly difficult to do and will not be achieved overnight. It will need courage and conviction because we will feel vulnerable during the process and this, as I argued at the beginning, is perhaps the reason why most OR workers choose to adopt the present limited view of OR.

REFERENCES

1. ACKOFF RL (1962) *Scientific Method: Optimising Applied Research Decisions*. John Wiley, New York.
2. BISHOP BC (1972) A contribution to a discussion on the methodology of OR. *Op. Res. Q.* **23**(3), 251–260.
3. CHURCHMAN CW (1972) *Design of Inquiring Systems: Basic Concepts in Systems Analysis*. Basic Books, New York.
4. DANDO MR & SHARP RG (1977) Technicians, advisers—and all-rounders in operational research. *Omega* **5**(5), 506–508.
5. DANDO MR & EDEN C (1977) Reflections on Operational Research—A Report from the EURO II Congress. *Omega* **5**(3), 255–270.
6. DANDO MR & SHARP RG (1977) *Operational Research in the UK in* 1977: *The Causes and Consequences of a Myth*? Presented at the session 'Managing OR Effectively' at the UK OR Society Annual Conference, Oxford.
7. EILON S (1977) Technician or adviser? *Omega* **5**(1), 1–6.
8. LAW B (1975) *The Concept of Counselling* (2). Papers prepared by a working party of the Standing Conference for the Advancement of Counselling. Bedford Square Press, London.
9. SHAW J (1977) *Basic Counselling: An Introduction for Managers, Supervisors, Teachers, Voluntary Social Workers and Others*. Vernon Scott, Stockport, UK.
10. VAUGHAN TD (Ed.) (1975) *Concepts of Counselling*: Papers prepared by a working party of the Standing Conference for the Advancement of Counselling. Bedford Square Press, London.
11. WHITE DJ (1975) *Decision Methodology: A Formalisation of the OR Process*. John Wiley, New York.

DAVID G SMITH

41, *Tutsham Way*
Paddock Wood
Kent TN12 6UA
UK

Vol. 6
1978

BLANK
PAGE

Vol.
197

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 211–225
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0211$02.00/0

Postwar Structural and Technological Changes in the American Economy

ROGER H BEZDEK

United States Department of Energy, Washington, D.C., USA

(*Received September 1975; in revised form November 1977*)

This paper utilizes input–output techniques to analyze postwar technological and structural change in the American economy, focusing on the subperiods 1947–1958, 1958–1963 and 1963–1966. The relationship between technological change, changes in input–output relationships and structural change is discussed. Postwar changes in total output requirements as measured by inverse coefficients and in intermediate output requirements as measured by inverse and by direct coefficients are summarized, and the variability in individual input–output coefficients is analyzed. It is found that postwar changes in the structure of the American economy have been highly uneven, unpatterned and erratic and that the accurate forecasting of these changes is extremely difficult given the data which are presently available. The indications of these findings for interindustry research are briefly discussed.

Vol. 6
1978

I. INTRODUCTION

INPUT–OUTPUT analysis is a technique for measuring and analyzing the interdependence of all industries in the economy. An input–output table traces the origin of all inputs and the destination of all the various outputs of all industries in the economy, and a simple hypothetical input–output table is illustrated in Table 1. This table depicts an economy consisting of only three intermediate sectors: Agriculture, Manufacturing, and Services. Reading across the row of one of these sectors, the middle figure in each cell shows the distribution of the output of the sector to the other (intermediate) processing sectors of the economy, to itself, and to the finished products comprising final demand (gross national product). For example, Table 1 indicates that Manufacturing sold 40 units of output to Agriculture, consumed 25 units of output itself, sold 90 units to Services, and delivered the remaining 80 units of its output to final consumers. Reading down a column of this table shows the sources of the inputs required by each sector. Reading down the Manufacturing column we find that this sector purchased 60 units of output from Agriculture, 25 units from itself, 70 units from Services, and 80 units from value added-wages, profits depreciation allowances, and taxes. US input–output data are integrated with the National Income and Product Accounts, and the sum of the value added components and the sum of the final demand components are both identical to gross national product (GNP).

The direct input-output coefficients given to the left in each cell are the ratio of the input shown to the total output of the sector in whose column it appears, and these coefficients tell, proportionately, how much a sector must purchase directly from every other sector to produce one unit of output. These direct purchases, however, give rise to subsequent rounds of indirect requirements. For example, steel mills require electricity to produce steel. But an electric utility requires turbines from a factory to produce electricity. The factory requires steel from steel mills to produce turbines, and the steel mill requires more electricity, ... and so on. The sum of the direct requirements plus the indirect requirements represents the total output requirements from a sector necessary to deliver one unit of output to final demand. Input–output techniques allow us to compute the direct as well as the indirect production requirements of an economy, and these total requirements are represented by the 'inverse'

TABLE 1. HYPOTHETICAL INPUT–OUTPUT TABLE

	Agriculture	Manufacturing	Services	Final demand	Gross output
Agriculture	0.30 *60* 1.23	0.26 *60* 0.71	0.08 *20* 0.13	*60*	200
Manufacturing	0.20 *40* 0.42	0.11 *25* 1.41	0.38 *90* 0.55	*80*	235
Services	0.05 *10* 0.19	0.30 *70* 0.15	0.23 *55* 1.38	*105*	240
Value added	0.45 *90*	0.33 *80*	0.31 *75*	*245*	

coefficients given to the right in each cell in Table 1.

In recent years input–output analysis has been applied to an increasingly wide range of problem areas. Input–output models have been integrated with econometric quarterly and long-run forecasting models, various types of interindustry manpower models have been developed, and input–output techniques are also being applied to energy-related problem areas. Since 1972 most macroeconometric models have met with somewhat less than unqualified success in forecasting inflationary trends in the American economy and Lawrence Klein has recently suggested that the input–output approach offers the best hope for explaining relative price movements in the future.[1] But despite this increasing interest in interindustry analysis by economists and others, relatively little is known about structural change and the stability of input–output coefficients for the American economy in the past three decades. Do input coefficients for most industries tend to be relatively stable over time? Can the interindustry structure of one year be used to approximate that of a different year? On the basis of the empirical data available, can methods be devised for predicting the level and direction of changes in detailed industry output requirements? Here we examine these and related questions by analyzing the structural change which occurred in the US economy between 1947 and 1966, focusing on the subperiods 1947–1958, 1958–1963 and 1963–1966.

[1] See Klein [6], pp. 12–13.

II. STRUCTURAL CHANGE AND CHANGES IN INPUT–OUTPUT COEFFICIENTS

Here we utilize the static open input–output model, which may defined as:

$$\mathbf{x} = (I - A)^{-1}\mathbf{y} \qquad (1)$$

where: $\mathbf{x}$ is an n-order total output vector, $\mathbf{y}$ is an n-order final demand vector, $(I - A)^{-1}$ is an n-order Leontief inverse matrix whose coefficients a^*_{ij} indicate the total output requirements generated from industry i by industry j per dollar delivered to final demand.

In this model changes in final demand are exogenous to the system and arise from changes in the level and composition of per-

TABLE 2. CHANGE OVER TIME IN TOTAL DOMESTIC OUTPUTS REQUIRED TO PRODUCE 1958 FINAL DEMAND

Industry number and title	Derived total domestic output[a]				Average annual rate of change		
	1947 Inverse	1958 Inverse	1963 Inverse	1966 Inverse	1947 to 1958	1958 to 1963	1963 to 1966
1 Livestock	29,103	26,091	26,932	25,604	−1.0	0.6	−1.7
2 Other ag. prod.	25,179	23,043	22,681	21,832	−0.8	−0.3	−1.3
3 Forestry and fishery products	965	1,149	854	816	1.6	−5.8	−1.5
4 Ag., forestry & fisheries serv.	2,156	1,564	1,463	1,401	−2.9	−1.3	−1.4
5 Iron ores mining	875	796	662	708	−0.9	−3.6	2.3
6 Nonferrous metal ores	1,180	1,031	782	724	−1.2	5.4	−2.5
7 Coal mining	4,557	2,749	2,384	2,357	−4.5	−2.8	−0.4
8 Crude petroleum and natural gas	9,989	9,669	9,259	8,578	−0.3	−0.9	−2.5
9 Stone & clay mining	1,227	1,508	1,555	1,579	1.0	0.6	0.5
10 Chem. & fertilizer mining	394	483	478	491	1.9	−0.2	0.9
11 New construction	52,416	52,416	52,416	52,416	—	—	—
12 Maintenance & repair construction	22,495	16,875	15,506	13,871	−2.5	−1.7	−3.6
14 Food & kindred prod.	61,812	63,865	64,213	62,813	0.3	0.1	−0.7
15 Tobacco manufactures	6,380	5,918	6,292	6,165	−0.7	1.2	−0.7
16, 17, 19 Textiles	13,525	15,094	14,971	14,830	1.0	−0.2	−0.3
18 Apparel	13,578	14,263	14,542	14,560	0.4	0.4	0.0

TABLE 2.—continued

Industry number and title	Derived total domestic output[a]				Average annual rate of change		
	1947 Inverse	1958 Inverse	1963 Inverse	1966 Inverse	1947 to 1958	1958 to 1963	1963 to 1966
20 Lumber & wood prod. (ex. cont.)	9,326	7,928	8,222	7,313	−1.5	0.7	−3.8
21 Wooden containers	815	445	334	337	−5.4	−5.6	0.3
22 Household furniture	3,360	3,300	3,207	3,154	−0.2	−0.6	−0.6
23 Other furniture	1,602	1,495	1,410	1,421	−0.6	−1.2	0.3
24, 25 Paper	12,718	13,167	13,100	12,908	0.3	−0.1	−0.5
26 Printing & publishing	13,060	12,613	11,750	11,155	−0.3	−1.4	−1.7
27 Chem. & selected chem. prod.	8,879	11,770	13,078	13,640	0.2	2.1	1.4
28 Plastic & synthetic material	2,387	4,233	5,289	6,300	5.3	4.6	6.0
29 Drugs, cleaning & toilet prep.	5,845	6,592	6,621	6,674	1.1	0.1	0.3
30 Paints & allied Prod.	2,141	1,874	1,921	1,864	−1.2	0.5	−1.0
31 Petroleum refining	17,276	17,343	18,214	17,566	0.0	1.0	−1.2
32 Rubber & misc. plastics prod.	6,021	6,842	8,074	8,446	1.2	3.4	1.5
33 Leather tanning & industrial leather prod.	1,043	881	801	746	−1.5	−1.9	−2.4
34 Footwear etc.	3,051	3,104	3,015	3,021	0.2	−0.6	0.1
35 Glass & glass prod	2,291	2,145	2,092	2,092	−0.6	−0.5	0.0
36 Stone & clay prod.	6,555	7,508	7,851	8,104	1.2	0.9	1.1
37 Primary iron & steel mfg.	26,087	19,107	18,654	19,024	−2.8	−0.5	0.7
38 Primary nonferrous metals mfg.	10,420	9,189	10,127	11,085	−1.1	2.0	3.1
39 Metal containers	1,949	2,095	2,010	2,075	0.7	−0.8	1.1
40 Heating, plumbing, etc.	8,062	8,035	7,960	8,783	0.0	−0.2	3.3
41 Screw mach. prod.	4,588	3,686	3,248	3,362	−2.0	−2.5	1.2
42 Other fab. metal prod.	5,959	6,440	6,649	6,899	0.7	0.6	1.2
43 Engines & turbines	2,264	2,200	2,237	2,324	−0.3	0.3	1.3
44 Farm mach. & equip.	2,320	2,439	2,370	2,394	0.5	−0.6	0.3
45 Construction machinery	3,253	3,084	2,960	3,013	−0.5	−0.8	0.6
46 Materials handling mach.	1,170	1,081	1,199	1,287	−0.7	2.1	2.4
47 Metalworking mach.	3,571	3,629	3,480	3,769	0.1	−0.8	2.7
48 Special industry mach.	2,632	2,509	2,565	2,602	−0.4	0.4	0.5
49 General industrial mach.	3,935	3,744	3,978	4,103	−0.5	1.2	1.0
50 Machine shop prod.	817	1,587	1,710	1,645	6.2	1.5	−1.3
51 Off., computing & acct. mach.	1,508	2,217	2,159	2,629	3.6	−0.5	6.8
52 Service ind. machines	2,230	2,249	2,727	3,047	0.1	3.9	3.8
53 Elec. transmission & dist. equip.	4,333	5,103	5,038	5,283	1.5	−0.3	1.6
54 Household appliances	3,161	3,594	3,928	3,945	1.2	1.8	0.1
55 Elect. light	2,287	2,287	2,318	2,390	0.0	0.3	1.0
56, 57 Comm. and electronic equip.	6,298	8,657	9,453	10,461	2.9	1.8	3.4
58 Misc. select. mach.	1,697	1,534	1,559	1,484	−0.9	0.3	−1.6
59 Motor vehicles	22,760	22,836	22,631	21,799	0.0	−0.2	−1.2
60, 13 Aircraft & ordinance	15,896	17,349	16,668	16,161	0.8	−0.8	−1.0
61 Other trans. equip.	3,839	3,723	3,628	3,669	−0.3	−0.5	0.4
62 Prof., scien. and control	2,939	3,498	3,355	3,369	1.6	−0.8	0.1
63 Optical, ophthalmic, & photo	1,367	1,542	1,728	1,796	1.1	2.3	1.3
64 Misc. mfg.	5,068	5,306	5,323	5,166	0.4	0.1	−1.0
65 Transportation	38,515	33,290	31,641	31,481	−1.4	−1.0	−0.2
66 Communications	7,550	9,292	9,876	9,998	1.9	1.2	0.4
67 Radio & TV broadcasting	779	1,549	1,583	1,660	6.4	0.4	1.6
68 Elect. gas, etc.	15,710	20,253	21,613	21,467	2.3	1.3	−0.2
69 Wholesale & retail trade	92,031	95,250	93,311	92,747	0.3	−0.4	−0.2
70 Finance and insurance	27,077	26,464	23,439	22,591	−0.2	−2.4	−1.2
71 Real estate & rental	60,249	61,934	62,729	62,157	0.3	0.3	−0.3
72 Hotel & lodging places	13,389	12,171	11,841	11,632	−0.9	−0.6	−0.6
73 Business services	18,906	24,441	24,941	26,133	2.4	0.4	1.6
74 Research & development	5,185	5,339	—	—	0.3	—	—
75 Auto. repair & services	9,079	7,913	7,637	7,576	−1.3	−0.7	−0.3
76 Amusements	4,781	5,620	5,327	5,378	1.5	−1.1	0.3
77 Medical, educ. serv., & non-profit organizations	23,474	22,703	22,639	22,656	−0.8	−0.1	0.0
81 Business, travel, etc.	7,132	6,451	5,682	5,647	−0.9	−2.5	−0.2
Total	816,468	819,144	811,890	808,176	96.4	90.6	93.4
Average[b]					1.3	1.3	1.3

[a] Millions of 1958 dollars.
[b] Averages computed ignoring signs.

sonal consumption expenditures, changes in investment spending, shifts in government programs, and related factors. These exogenous final demand changes result in associated changes in total output requirements from individual industries. However, changes in industry output requirements can also result from changes in the direct and indirect requirements of production—changes in input–output coefficients. Structural change can refer to either changes over time in industry output requirements caused by changes in both final demand and input–output coefficients or to changes only in the input–output relationships themselves. The two types of changes are related but not identical. Here we use the latter interpretation of structural change, changes over time in interindustry relationships only.

Changes in input–output relationships are often referred to as technological changes. However, while technological change—change in the physical requirements for the goods and services used in producing a specified combination of commodities—is an important and consistent cause of changes in interindustry relationships, it is important to note that coefficient changes can be caused by a number of other factors as well. When the economy is divided into 80 industries, an individual industry cannot represent a single or even a homogeneous set of commodities. Thus, shifts over time in the individual industry composition of the broad input–output industry—shifts in product mix—will cause associated shifts in the coefficients for that industry. A third factor causing coefficient change is the divergence of actual technical relationships from a linear-homogeneous production function and the proportional relationships which it assumes. That is, the coefficients for one year may differ from those of another merely because the scale of preparation or degree of utilization has shifted. Specific conventions concerning the handling of secondary products, dummy industries, and imports in input–output tables can influence the stability of coefficients over time. The level of aggregation can also affect the stability of individual coefficients. On the one hand more aggregate tables may possess unstable coefficients due to the problem of changing product mix, while, on the other hand aggregation may contribute to coefficient stability by cancelling out the impact of substitutions among related materials. Finally, changes in technical coefficients may reflect nothing more than random factors: data composition, sampling procedures, statistical conventions, and so forth.[2] Accordingly, here we do not refer to either structural change or changes in input–output coefficients as technological change, but merely not the close interdependence of these concepts.

III. DATA SOURCES AND CONVENTIONS

For our study we required input–output data for the years, 1947, 1958, 1963 and 1966. Current dollar total industry base data are influenced by changes in prices, transferred imports, and by the patterns of secondary output transferred to industries for which such output is primary. To obtain a more accurate picture of postwar structural change it was necessary to eliminate the influence of these factors by utilizing constant (1958) dollar primary output base data. Primary output includes all output of the product in question wherever produced and includes primary establishment output as well as the 'transfers in' of the product in question produced as a secondary product in other establishments.[3]

The 1947 input–output study was conducted in the early 1950's by the Bureau of Labor Statistics, but in 1970, the 1947 interindustry data were reworked and deflated by BEA staff to integrate them into the National Income and Product Accounts and to make them statistically and conceptually consistent with later BEA interindustry studies.[4] Due to the extensive revisions between the 1945 and the 1957 Standard Industrialization Classification systems for some industries, the reconciliation of

[2] For a more complete discussion of these problems see Vaccara [15], pp. 3–8.

[3] The direct coefficients used in our analysis were computed excluding transfers in; that is, the base used to divide each column of the transactions matrix did not include any transfers in that column. The row output totals used in conjunction with these coefficients excluded all transfers out. The inverse matrices and the data used in conjunction with them were not developed according to these latter concepts, but were developed only on a constant dollar domestic output base. These conventions differed somewhat from those adhered to in the earlier Vaccara studies [14], [15].

[4] See [12].

the 1947 data was not entirely adequate. Thus in all of the analyses conducted here the following industries have been combined to avoid errors resulting from inaccurate classification: 13 and 60; 16, 17, and 19; 24 and 25; 56 and 57, (see Tables 2, 3 and 4).

The source of the 1958 data was the BEA 1958 interindustry table which, though originally published on a total base, was also available on a domestic output base.[5] A number of special adjustments were made in the 1958 data to make them consistent with the conventions followed in the 1963 and 1966 interindustry studies. The 1963 current dollar data were available from the 1963 input–output study and the 1966 current dollar data were available from the 1966 interindustry update study. The 1963 data and the 1966 data were deflated to 1958 prices by BEA staff.[6]

IV. POSTWAR CHANGES IN INDUSTRY OUTPUT REQUIREMENTS

A straightforward way of summarizing the impacts on industry output requirements of changes over time in interindustry relationships is to compare the total outputs required from each industry to produce a given (1958) final demand bill of goods with the input–output structures of 1947, 1958, 1963 and 1966. Accordingly, here we multiplied the domestic inverse matrices from each of these years by the 1958 final demand vector to estimate the outputs required from each industry to produce the 1958 bill of goods with input–output structures of the four different years:

$$\mathbf{x}_t = (I - A_t)^{-1}\mathbf{y}_{1958} \qquad (2)$$

where $\mathbf{y}_{1958}$ is the 1958 domestic final demand vector, $(I - A_t)^{-1}$ is the constant dollar domestic Leontief inverse matrix for years $t = 1947, 1958, 1963, 1966$, $\mathbf{x}_t$ is the derived output requirements vector for the years $t = 1947, 1958, 1963, 1966$.

The results of this analysis are summarized in Table 2. This table indicates that to produce the 1958 bill of goods with 1958 technology required 0.4% more output as compared to 1947 technology, that to produce 1958 final demand with 1963 technology required a 0.1% less output than with 1958 technology, and to produce the 1958 bill of goods with 1966 technology required 1.1% less output as compared to 1963 technology. However, a more meaningful way of comparing the differences in output is to compute the average change in industry outputs ignoring signs.

To standardize for the different lengths of the time periods involved we computed the average annual rates of change in total output requirements over each subperiod. Table 2 indicates that while the average annual rates of change in total output requirements varied

[5] The 1958 study was published in Goldman, Marimont and Vaccara [5].

[6] The 1963 and the 1966 studies were published in [11] and [13]. The deflation of the 1963 data is discussed in Frazier [4]; the deflation of the 1966 data is discussed in Bezdek, Frazier and Wendling [1].

TABLE 3. CHANGE OVER TIME IN INTERMEDIATE DOMESTIC OUTPUTS REQUIRED TO PRODUCE 1958 FINAL DEMAND

Industry number and title	Derived intermediate domestic output[a]				Average annual rate of change		
	1947 Inverse	1958 Inverse	1963 Inverse	1966 Inverse	1947 to 1958	1958 to 1963	1963 to 1966
1 Livestock	26,576	23,564	24,403	23,075	−1.1	0.7	−1.8
2 Other ag. prod.	19,760	17,624	16,640	15,791	−1.0	−1.1	−1.7
3 Forestry & fishery prod.	1,073	1,247	963	925	1.4	−5.2	−1.3
4 Ag., forestry & fisheries serv.	2,156	1,564	1,463	1,041	−2.9	−1.3	−1.4
5 Iron ores mining	1,306	1,227	1,091	1,137	−0.6	−2.3	1.4
6 Nonferrous metal ores	1,305	1,156	906	828	−1.1	−4.8	−2.2
7 Coal mining	3,929	2,121	1,755	1,728	−5.5	−3.7	−0.5
8 Crude petroleum & natural gas	11,185	10,865	10,454	9,773	−0.3	−0.8	−2.2
9 Stone & clay mining	1,302	1,583	1,629	1,653	1.8	0.6	0.5
10 Chem. & fertilizer mining	396	485	480	492	1.9	−0.2	0.9
11 New construction	0	0	0	0	—	—	—
12 Maintenance & repair const.	18,075	12,455	11,086	9,451	−3.3	−2.3	−5.2
14 Food & kindred prod.	15,484	17,597	17,680	16,280	1.2	0.1	−2.7
15 Tobacco manufactures	1,746	1,284	1,659	1,532	−2.7	5.3	−2.6
16, 17, 19 Textiles	11,199	12,768	12,644	12,504	1.2	−0.2	−0.4
18 Apparel	2,297	2,982	3,260	3,278	2.4	1.8	0.2

Vol. 6 1978

TABLE 3.—continued

Industry number and title	Derived intermediate domestic output[a] 1947 Inverse	1958 Inverse	1963 Inverse	1966 Inverse	Average annual rate of change 1947 to 1958	1958 to 1963	1963 to 1966
20 Lumber & wood prod. (ex. cont.)	9,515	8,117	8,411	7,502	−1.5	0.7	−3.7
21 Wooden containers	825	455	344	347	−5.3	−5.5	0.3
22 Household furniture	726	666	573	520	−0.8	3.0	−3.2
23 Other furniture	503	396	312	323	−2.2	−4.7	1.2
24, 25 Paper	12,409	12,908	12,841	12,649	0.4	−0.1	−0.5
26 Printing and publishing	10,167	9,840	8,976	8,381	−0.3	−1.8	−2.3
27 Chem. & selected chem. prod.	7,367	10,260	11,484	12,046	3.1	2.3	1.6
28 Plastic and synthetic material	2,115	3,961	5,011	6,022	5.9	4.8	6.3
29 Drugs, cleaning & toilet prep.	1,487	2,234	2,246	2,299	3.8	0.1	0.8
30 Paints & allied prod.	2,099	1,832	1,879	1,822	−1.2	0.5	−1.0
31 Petroleum refining	9,095	9,162	10,013	9,365	0.1	1.8	−2.2
32 Rubber & misc. plastics prod.	4,321	5,143	6,362	6,34	1.6	4.3	1.9
33 Leather tanning & industrial leather products	1,054	892	813	758	−1.5	−1.8	−2.3
34 Footwear etc.	362	415	325	331	1.2	−4.8	0.6
35 Glass and glass prod.	2,157	2,011	1,957	1,957	−0.6	−0.5	0.0
36 Stone and clay prod.	6,306	7,259	7,601	7,854	1.3	0.9	1.1
37 Primary iron & steel mfg.	25,862	18,882	18,425	18,795	−2.8	−0.5	0.7
38 Primary nonferrous metals mfg.	10,690	9,459	10,385	11,343	−1.1	1.9	3.0
39 Metal containers	1,883	2,029	1,943	2,008	0.7	−0.9	1.1
40 Heating, plumbing, etc.	7,146	7,119	7,029	7,852	0.0	−0.3	3.8
41 Screw mach. prod.	4,308	3,406	2,966	3,080	−2.2	−2.7	1.3
42 Other fab. metal prod.	5,156	5,637	5,829	6,079	0.8	0.7	1.4
43 Engines & turbines	1,175	1,111	1,100	1,187	−0.5	−0.2	2.6
44 Farm mach. & equip.	580	699	618	642	1.7	−2.4	1.2
45 Construction machinery	1,198	1,029	901	954	−1.4	−2.6	1.9
46 Materials handling mach.	595	506	621	709	−1.5	4.2	4.5
47 Metalworking mach.	2,048	2,106	1,855	2,144	0.3	−2.5	4.9
48 Special industry mach.	854	731	783	820	−1.4	1.4	1.6
49 General industrial mach.	2,496	2,305	2,534	2,659	−0.7	1.9	1.6
50 Machine shop prod.	753	1,523	1,644	1,579	6.6	1.5	−1.3
51 Off., computing and acct. mach.	190	900	831	1,301	15.2	−1.6	16.1
52 Service ind. machines	847	866	1,657	0.2	9.1	7.4	
53 Elec. transmission & dist. equip.	2,423	3,193	2,959	3,204	2.5	−1.5	2.7
54 Household appliances	532	965	1,149	1,166	5.6	3.6	0.5
55 Elect. light	1,899	1,899	1,859	1,931	0.0	−0.4	1.3
56, 57 Comm. and Electronic equip.	1,928	4,287	4,459	5,577	7.5	1.3	6.9
58 Misc. elect. mach.	1,204	1,041	1,044	969	−1.3	0.1	−2.4
59 Motor vehicles	9,484	9,560	9,221	8,389	0.1	−0.7	−3.1
60, 13 Aircraft & Ordinances	6,253	7,706	4,419	1.9	−10.5	−4.0	
61 Other trans. equip.	1,078	962	866	907	−1.0	−2.1	1.6
62 Prof., scien. and control	1,361	1,920	1,668	1,682	3.2	−2.8	0.3
63 Optical, ophthalmic, & photo	564	740	895	963	2.5	3.9	2.4
64 Misc. mfg.	2,106	2,375	2,386	2,229	1.1	0.1	−2.2
65 Transportation	25,034	19,809	17,981	17,821	−2.2	−1.9	−0.3
66 Communications	2,856	4,598	5,182	5,304	4.4	2.4	0.8
67 Radio & TV broadcasting	770	1,540	1,574	1,651	6.5	0.4	1.6
68 Elect. gas, etc.	6,817	11,360	12,721	12,575	4.8	2.3	−0.4
69 Wholesale & retail trade	23,772	26,991	24,973	24,408	1.2	−1.5	−0.8
70 Finance and insurance	15,003	14,390	11,365	10,518	−0.4	−4.6	−2.6
71 Real estate & rental	18,477	20,162	20,958	20,386	0.8	0.8	−0.9
72 Hotel & lodging places	3,601	2,383	2,053	1,844	−3.7	−2.9	−3.5
73 Business services	15,722	21,257	21,385	22,577	2.8	0.1	1.8
74 Research & development	—	—	—	—	—	—	—
75 Auto. repair & services	4,480	3,314	3,038	2,977	−2.7	−1.7	−0.7
76 Amusements	1,265	2,104	1,810	1,861	4.7	−3.0	0.9
77 Medical, educ. serv., & nonprofit organizations	2,594	1,823	1,221	1,237	−3.2	−7.7	0.4
81 Business, travel, etc.	7,292	6,611	5,842	5,807	−0.9	−2.4	−0.2
Total	410,593	413,381	405,200	401,485	161.3	161.1	150.7
Average[b]					2.3	2.3	2.1

[a] Millions of 1958 dollars.
[b] Averages computed ignoring signs.

between 1% and 2.6%, the rates of change experienced by individual industries varied much more. As is to be expected, there are some industries which during each subperiod experienced average annual rates of growth in output requirements well in excess of the overall average and there are other industries which consistently exhibited an average annual rate of growth in output requirements well below the overall average.

However, the interpretation of the above findings can be misleading, for industries which sell primarily to final demand will tend to show smaller percent changes over time in total output requirements than industries which sell primarily to the intermediate processing sectors. For this reason a more useful measure of output change can be gained by looking at the impacts on intermediate output requirements. This permits us to abstract from the fact that for many industries a sizable portion of total output is determined more by final demand than by input–output coefficients, and is thus largely held constant over time by the method of analysis employed above.[7] Table 3 compares the changes in intermediate output requirements needed to produce the 1958 bill of goods with different technologies over the period 1947–1966. These estimates were computed by subtracting final demand requirements from the output requirements generated by use of the different inverse matrices:

$$\mathbf{Q}_t^z = (I - A_t)^{-1}\,\mathbf{x}_{1958} - \mathbf{y}_{1958} \qquad (3)$$

where: $\mathbf{Q}_t^z$ is the derived intermediate output requirements vector for the years $t = 1947, 1958, 1963, 1966$.

Comparison of Tables 2 and 3 indicates that in each subperiod the average change in intermediate output requirements was considerably larger than was the average change in total output requirements. In the period 1947–1958 the annual rate of change in intermediate output requirements averaged 2.3% whereas the change in total output requirements averaged 1.3%; between 1958 and 1963 the change in intermediate output requirements averaged 2.3% whereas the change in total output requirements averaged 1.3%; and between 1963 and 1966 the change in intermediate output requirements averaged 2.1% as compared to 1.3% for total output requirements. Thus, even after adjusting for the differences in the lengths of the subperiods involved, it is clear that the rate of change of intermediate output requirements differed appreciably between subperiods. Aside from this it is difficult to make any generalizations about the differences in the rates of output change in each subperiod. At the very least, though, it appears that the slowdown in the rate of change in intermediate output requirements noted earlier by Beatrice Vaccara for the period 1958–1961 did not continue through 1963.[8] Thus, while the direction and magnitude of the rate of change in intermediate output requirements for many industries varied between 1958 and 1963, there was no decline in the overall rate of change in output requirements between 1958 and 1963 as compared to that which occurred between 1947 and 1958. In fact, nearly half (33) of the industries experienced a higher annual rate of change in intermediate output requirements between 1958 and 1963 than they did between 1947 and 1958.

Even focusing on the data derived only from the benchmark tables (1947, 1958, 1963) it is clear that a major finding which emerges here is the highly erratic and uneven nature of the average annual rates of intermediate output change experienced by most industries between 1958 and 1963 and those registered between 1947 and 1958. Not only does the magnitude of the average rate of change in intermediate output for most industries vary widely, but for more than one out of three industries even the direction of the rate of output change differed between 1947–1958 and 1958–1963. Viewed from a different perspective, we can ask ourselves for which industries would the average annual rate of change in intermediate output requirements between 1947 and 1958 have provided a relatively accurate guide in forecasting the average annual rate of output change through 1963. The erratic and uneven nature of these rates of change is again apparent, for there were only 11 out of 71 industries which experienced approximately the same rate and direction of intermediate output change between 1958 and 1963 as between 1947 and 1958.

For an analysis of structural change which attempts to identify the underlying causes of these changes it is important to measure separ-

[7] See Vaccara [14], p. 12.
[8] See Vaccara [14], p. 13.

TABLE 4. CHANGES OVER TIME IN PRIMARY INTERMEDIATE OUTPUTS REQUIRED TO PRODUCE 1958 FINAL DEMAND[a]

Industry number and title	Derived intermediate domestic output (millions of 1958 dollars)				Average annual rate of charge		
	1947 direct coef.	1958 direct coef.	1963 direct coef.	1966 direct coef.	1947 to 1958	1958 to 1963	1963 to 1966
1 Livestock	23,685	21,288	22,457	21,788	−1.0	1.1	−1.0
2 Other ag. prod.	15,365	15,250	14,770	14,906	−0.1	−0.1	0.3
3 Forestry & fishery prod.	992	1,240	933	948	2.1	−5.5	0.5
4 Ag., forestry & fisheries serv.	2,044	1,558	1,482	1,472	−2.4	−1.0	−0.2
5 Iron ores mining	972	1,197	1,165	1,129	0.9	−1.6	0.7
6 Nonferrous metal ores	1,152	1,127	856	759	−0.2	−5.4	−3.9
7 Coal mining	3,669	2,102	1,763	1,720	−4.9	−3.5	−0.8
8 Crude petroleum & natural gas	11,045	10,330	9,463	9,091	−0.6	−1.7	−1.3
9 Stone & clay mining	1,305	1,506	1,547	1,580	1.3	0.5	0.7
10 Chem. & fertilizer mining	399	325	287	287	−1.9	−2.5	−0.1
11 New construction	0	0	0	0	—	—	—
12 Maintenance & repair construction	18,609	12,455	10,924	9,414	−3.6	−2.6	−4.8
14 Food & kindred prod.	13,399	14,856	15,059	14,165	0.9	0.3	−2.0
15 Tobacco manufactures	1,451	1,133	1,420	1,348	−2.2	4.6	−1.7
16, 17, 19 Textiles	11,580	11,921	11,728	11,626	0.3	−0.3	−0.3
18 Apparel	2,351	2,813	3,000	3,028	1.6	1.3	0.3
20 Lumber & wood prod. (ex. cont.)	8,625	7,901	8,244	7,707	−0.8	0.9	−2.2
21 Wooden containers	796	423	320	325	−5.6	−5.4	0.5
22 Household furniture	682	541	457	407	−2.1	−3.3	−3.9
23 Other furniture	404	291	235	257	−2.9	−4.2	3.1
24, 25 Paper	13,012	12,518	12,435	12,414	−0.4	−0.1	−0.1
26 Printing & publishing	5,413	4,141	3,733	3,702	−2.4	−2.1	−0.3
27 Chem. & selected chem.	7,491	9,281	9,924	10,036	2.0	1.4	0.4
28 Plastic & synthetic material	2,105	3,602	4,266	5,047	5.0	3.4	5.8
29 Drugs, cleaning & toilet prep.	1,274	1,794	1,727	1,786	3.2	−0.8	1.1
30 Paints and allied prod.	1,745	1,719	1,832	1,884	−0.1	1.3	0.9
31 Petroleum refining	8,682	8,305	8,786	8,301	−0.4	1.1	−1.9
32 Rubber & misc. plastics prod.	4,104	4,674	5,859	6,129	1.2	4.6	1.5
33 Leather tanning and industrial leather products	1,031	881	827	610	−1.4	−1.3	−9.7
34 Footwear etc.	307	326	244	221	0.6	−5.7	−3.2
35 Glass & glass prod.	2,202	1,976	1,911	1,966	−1.0	−0.7	1.0
36 Stone & clay prod.	6,112	6,906	7,154	7,442	1.1	0.7	1.3
37 Primary iron & steel mfg.	23,529	17,672	15,542	18,092	−2.6	−0.2	1.0
38 Primary nonferrous metals mfg.	9,885	8,907	9,384	9,919	−0.9	1.1	1.9
39 Metal containers	1,845	1,974	1,841	1,930	0.6	−1.4	1.6
40 Heating, plumbing, etc.	6,083	6,512	6,589	7,463	0.6	0.2	4.2
41 Screw mach. prod.	3,948	2,991	2,537	2,661	−2.5	−3.2	1.6
42 Other fab. metal prod.	4,572	5,111	5,318	5,553	1.0	0.8	1.5
43 Engines & turbines	799	804	899	964	0.1	2.3	2.4
44 Farm mach. & equip.	260	434	365	359	4.8	−3.4	−0.5
45 Construction machinery	769	687	618	651	−1.0	−2.1	1.7
46 Materials handling mach.	421	378	484	528	−1.0	5.1	3.0
47 Metalworking mach.	1,781	1,772	1,565	1,756	−0.1	−2.5	3.9
48 Special industry mach.	536	438	452	449	−1.8	0.6	−0.2
49 General industrial mach.	1,948	1,919	2,081	2,119	−0.1	1.6	0.6
50 Machine shop prod.	687	1,376	1,489	1,418	6.5	1.6	−1.6
51 Off., computing & acct. mach.	144	328	454	436	7.8	6.7	−1.3
52 Service ind. machines	596	584	1,041	1,347	−0.2	12.3	9.0
53 Elec. transmission & dist. equip.	2,109	2,498	2,383	2,539	1.6	−0.9	2.1
54 Household appliances	137	447	515	448	11.3	2.9	−4.5
55 Elect. light.	1,683	1,631	1,674	1,768	−0.3	0.5	1.8
56, 57 Comm. and electronic equip.	1,930	3,056	3,359	44093	4.3	1.9	6.8
58 Misc. elect. mach.	1,054	841	868	778	−2.0	0.6	−3.6
59 Motor vehicles	8,678	8,386	8,446	7,838	−0.3	0.1	−2.5
60, 13 Aircraft & ordnance	1,673	3,488	2,861	2,787	6.9	−3.9	−0.9
61 Other trans. equip.	863	683	592	562	−2.1	−2.8	−1.7
62 Prof., scien. and control	1,155	1,413	1,249	1,251	1.9	−2.4	0.1
63 Optical, ophthalmic & photo	492	583	708	715	1.6	4.0	0.3
64 Misc. mfg.	1,508	1,771	1,757	1,622	1.5	−0.2	−2.6
65 Transportation	20,004	16,743	16,028	15,682	−1.6	−0.9	−0.7

TABLE 4.—continued

Industry number and title	Derived intermediate domestic output (millions of 1958 dollars)				Average annual rate of charge		
	1947 direct coef.	1958 direct coef.	1963 direct coef.	1966 direct coef.	1947 to 1958	1958 to 1963	1963 to 1966
66 Communications	2,896	4,193	4,893	5,063	3.4	3.1	1.1
67 Radio & TV broadcasting	8	8	20	20	0.0	20.5	−1.0
68 Elect. gas, etc.	7,413	11,302	12,389	12,250	3.9	1.9	−0.4
69 Wholesale & retail trade	21,977	25,734	24,052	23,650	1.5	−1.3	−0.6
70 Finance and insurance	14,234	13,620	11,140	10,366	−0.4	−3.9	−2.4
71 Real estate & rental	18,354	20,163	21,353	21,056	0.9	1.2	−0.5
72 Hotel & lodging places	2,139	1,063	1,023	913	−6.2	−0.8	−3.7
73 Business services	16,080	20,688	21,110	22,447	2.3	0.4	2.1
74 Research & development							
75 Auto, repair & services	4,298	3,244	3,021	3,018	−2.5	−1.4	0.0
76 Amusements	1,266	1,855	1,579	1,628	3.5	−3.2	1.0
77 Medical, educ. serv. & nonprofit organizations	1,497	1,224	1,173	1,216	−1.8	−0.8	1.2
81 Business, travel, etc.	7,503	6,611	5,897	5,841	−1.1	−2.3	−0.3
Total	368,757	367,512	365,506	364,691	149.7	176.1	133.5
Average[a]					2.1	2.5	1.9

[a] Averages computed ignoring signs.

ately the direct impact on an industry's output requirements of changes in interindustry structure. Table 4 shows the changes over time in derived primary intermediate output requirements for individual industries computed using direct rather than inverse coefficients. These data were derived by multiplying the actual 1958 domestic primary output for each industry in turn by the 1947, the 1958, the 1963, and the 1966 primary direct coefficient matrices (each in 1958 prices). These computations yielded the intermediate output that would have been required directly from each industry to produce the 1958 output levels with the 1947, the 1958, the 1963 and the 1966 input–output structures:

$$\mathbf{Q}_t = A_t \mathbf{x}_{1958} \tag{4}$$

where: $\mathbf{x}_{1958}$ is the 1958 domestic primary output vector, A_t is the constant dollar primary direct coefficient matrix for the years 1947, 1958, 1963, 1966. $\mathbf{Q}_t$ is the derived intermediate output requirements vector for the years t = 1947, 1958, 1963, 1966.

Comparison of Tables 3 and 4 indicates that for each subperiod between 1947 and 1966 the average annual rate of change of intermediate output requirements, in total, is similar irrespective whether it is derived using inverse coefficients or direct coefficients. Using inverse coefficients the average annual rates of change of total intermediate output were for the subperiods, 1947–1958, 1958–1963, and 1963–1966, respectively, 2.3%, 2.3%, and 2.1%. Using direct coefficients the average annual rates of intermediate output change for these same subperiods were 2.1%, 2.5%, and 1.5%. But while the averages for all industries combined were quite similar, this was not true for many individual industries. In each subperiod, there were many industries for which the implied average annual rates of change computed by the two methods differed considerably. The average annual rate of change in intermediate output requirements, ignoring signs, differed by one percent or more for 17 industries between 1947 and 1958, for 11 industries between 1958 and 1963 and for 21 industries between 1963 and 1966.

While the measures of intermediate output change based on inverse coefficients may not be an adequate guide to the degree of change over time in the various direct input coefficients, they can generally be regarded as a good guide to the direction of change in the direct input coefficients. Between 1947 and 1958 only 8 industries showed opposite direction of change for intermediate output requirements based on inverse as opposed to direct coefficients. Between 1958 and 1963 the direction in the rate of change of intermediate output requirements as indicated by inverse and by direct coefficients was in even closer agreement and differed in only 7 industries, while between 1963 and 1966 these measures differed

in direction in only 6 industries. For all three subperiods most of these cases of opposite direction of movement occurred in industries showing relatively small average annual rates of change based on both direct and inverse requirements.

From Table 4 we can again identify certain industries which between 1947 and 1966 consistently experienced rates of growth in intermediate output requirements (as measured by direct coefficients) which were above average and certain other industries which experienced rates of growth in intermediate output requirements which were consistently below average. It is interesting to determine whether those industries consistently identified as experienced high rates of growth in implied output requirements and those industries identified as experienced very low rates of implied changes in output requirements in Tables 2, 3, and 4 are the same in each case. That is, we wish to determine if using implied rates of change in total output requirements, implied rates of change in intermediate output requirements as measured by inverse coefficients, and implied rates of change in intermediate output requirements as measured primary direct coefficients we identify the same general types of industries as being affected the most, both positively and negatively, by postwar structural change. An examination of the tables mentioned above indicates that this is indeed the case. Of the nine or ten industries listed as experiencing the highest rates of implied growth in outputs in Tables 2, 3 and 4, the following seven industries appeared in each table: Plastics (28), Communications equipment and electronic components (56, 57) Chemicals (27), Rubber (32), Business services (73), Optical equipment (63) and Communications (66). Similarly, of the nine or ten industries shown as experiencing the lowest rates of implied growth in outputs in Tables 1, 2 and 3 the following seven appeared in each table: Maintenance and repair construction (12), Other furniture (23), Screw machine products (41), Wooden containers (21), Coal mining (7), Hotels and lodging places (72) and Nonferrous metals mining (6). It thus appears that between 1947 and 1966 the relative impact of structural change on the output requirements within different industries was similar irrespective of whether this change was measured as the impact on total output requirements using inverse coefficients, as the impact on intermediate output requirements using inverse coefficients or as the impact on primary intermediate output requirements using primary direct coefficients.

V. ANALYSIS OF VARIABILITY OF DIRECT COEFFICIENT CHANGES

Until now, our analysis has been concerned exclusively with average changes over time in the requirements for the output of individual industries generated using the 1958 final demands and input–output matrices from different years. To complete this study we turn next to an examination of the variability of the cell-by-cell coefficients for an industry to determine whether or not the input coefficients for the various direct customers of a given producing industry change over time in the same way as the overall average. This analysis relating to the variability of the individual primary direct input coefficients has to be confined to the two subperiods deliniated by benchmark years, 1947–1958 and 1958–1963. It would not be valid to include the 1963–1966 subperiod in this analysis since, in general, the procedures for updating the 1963 table to 1966 involved adjusting the benchmark coefficients of a given producing industry by the average change for that industry.

The results of this analysis for the periods 1947–1958, 1958–1963 are summarized in Table 5.[9] For any given producing industry (listed in column 1) the first entry in column 3 of this table is a weighted average of the percent change in the primary direct coefficients for all of its intermediate customers between 1947 and 1958. The next two entries indicate the percent change in the primary direct coefficients for the first and second largest customers (listed in column 2) of the given producing industries. The ranking of customer size is determined by the level of the dollar flows rather than by the size of the input coefficients.

[9] It should be noted that the data given in Table 4 for 1947–1958 are revised versions of those derived in the earlier Vaccara studies [14], [15] and were computed using total domestic output. As noted, all other computations here utilizing direct coefficient matrices employed domestic primary base data. The estimates in Table 4 are thus not comparable to those in Table 3, although the differences are quite small.

TABLE 5. VARIABILITY OF PRIMARY DIRECT COEFFICIENTS, 1947–58, 1958–63

Producing industry	Consuming industry	Percent change in primary		Direct coefficient		Percent of coefficients conforming		Producing industry	Consuming industry	Percent change in primary		Direct coefficient		Percent of coefficients conforming	
		1947–1958	1958–1963	1947–1958	1958–1963					1947–1958	1958–1963	1947–1958	1958–1963		
1	All	−9.2	+5.5	57	71	10	All	−21.9	−11.7	100	100				
	2[a]	−10.6	+1.4				1	−31.2	−6.4						
	1[b]	+24.3	+18.6				2	−20.0	−24.1						
2	All	−0.4	−3.1	71	75	12	All	−31.7	−12.3	84	86				
	1	+3.5	−4.6				1	−39.1	−11.7						
	2	−4.7	+3.8				2	−37.0	−20.7						
3	All	+26.3	−24.8	50	67	14	All	+11.7	+1.4	38	71				
	1	+56.1	−23.9				1	+25.5	−1.5						
	2	+20.1	−23.8				2	+30.9	−5.4						
4	All	−21.2	−4.9	67	33	15	All	−21.9	+25.3	100	100				
	1	−39.7	+1.6				1	−21.9	+25.3						
	2	+10.7	−25.7				2	−100.0	−100.0						
5	All	+22.9	−7.7	100	100	16,17,	All	+17.4	+3.1	38	69				
	1	+16.2	−7.4			19	1	+5.3	−7.2						
	2	+144.4	−25.3				2	−4.7	−7.8						
6	All	−2.6	24.1	67	100	18	All	+18.8	+6.7	100	50				
	1	−5.3	−26.2				1	+10.0	+5.7						
	2	+20.9	−9.9				2	−98.1	−43.6						
7	All	−41.9	−16.1	82	80	20	All	−8.5	+4.4	70	56				
	1	−48.5	−11.4				1	−21.3	−7.1						
	2	−1.6	−23.0				2	+55.7	+6.3						
8	All	−7.2	−8.4	50	33	21	All	−47.3	−24.4	83	100				
	1	−9.9	−13.0				1	−48.3	−13.4						
	2	+19.5	+22.9				2	−39.9	−23.0						
9	All	+14.4	+2.7	50	67	22	All	−22.6	−15.5	33	80				
	1	−16.7	−35.8				1	−56.7	−5.5						
	2	+68.5	+32.1				2	+21.6	−41.9						
23	All	−28.0	−19.4	50	0	33	All	−15.5	−6.2	50	67				
	1	−100.0	−23.2				1	−15.5	−14.0						
	2	−100.0	+8.4				2	−37.2	+37.2						
24,25	All	−5.0	+0.8	62	65	34	All	+4.8	−25.2	100	100				
	1	−2.7	+5.6				1	+27.3	−65.5						
	2	+1.7	+16.1				2	−76.0	−56.6						
26	All	−24.2	−11.7	89	83	35	All	−10.5	−3.3	73	69				
	1	+0.4	−13.6				1	−5.1	+16.5						
	2	−32.0	−33.2				2	−47.5	−9.5						
27	All	+21.1	+7.0	56	69	36	All	+12.6	+3.6	50	35				
	1	+24.0	−7.3				1	+14.8	+21.9						
	2	23.6	+12.2				2	+48.0	−7.5						
28	All	+68.6	+18.4	67	39	37	All	−25.3	−0.7	84	54				
	1	+92.6	+8.4				1	−3.1	o2.0						
	2	+127.4	+27.8				2	−18.1	−17.0						
29	All	+39.5	−3.8	67	75	38	All	−11.1	+5.4	54	51				
	1	+10.5	−11.2				1	+22.8	+12.2						
	2	+196.8	−4.3				2	−37.5	+21.9						
30	All	−1.9	+6.6	67	73	39	All	+7.2	−6.7	33	40				
	1	+36.8	−11.3				1	+48.8	−13.2						
	2	−29.6	+33.3				2	−47.9	−11.2						
31	All	−4.6	+5.8	77	34	40	All	+6.9	+1.2	22	63				
	1	−24.3	−3.8				1	+32.0	+7.1						
	2	+41.3	+10.6				2	−24.7	−39.5						
32	All	+13.0	+25.4	50	59	41	All	−25.0	−15.2	92	74				
	1	−41.9	−11.4				1	−30.3	+1.3						
	2	+2.8	+59.8				2	−8.9	−44.9						
42	All	+11.1	+4.1	33	62	51	All	+127.8	+39.2	100	50				
	1	+3.9	−3.4				1	+323.6	+100.2						
	2	+32.1	−22.3				2	−84.2	−43.4						

TABLE 5.—continued

Producing industry	Consuming industry	Percent change in primary 1947–1958	Direct coefficient 1958–1963	Percent of coefficients conforming 1947–1958	Percent of coefficients conforming 1958–1963	Producing industry	Consuming industry	Percent change in primary 1947–1958	Direct coefficient 1958–1963	Percent of coefficients conforming 1947–1958	Percent of coefficients conforming 1958–1963
43	All	+1.4	+11.8	11	33	52	All	−3.5	+78.2	7.5	67
	1	−17.9	−8.6				1	+4.7	+105.1		
	2	−10.2	+1.8				2	26.7	+32.7		
44	All	+66.3	−15.9	0	67	53	All	+16.7	−4.6		57
	1	−28.2	−6.4				1	+23.3	−29.2	47	
	2	+122.8	−57.5				2	−45.8	+0.2		
45	All	−10.8	−10.0	38	100	54	All	+226.3	+15.1	50	100
	1	−38.9	−25.6				1	+409.3	−12.8		
	2	−82.4	+13.5				2	+67.2	+14.7		
46	All	−10.8	+28.0	67	67	55	All	−3.5	+2.7	86	38
	1	+24.1	−9.6				1	+24.0	+20.8		
	2	+41.6	+76.0				2	−47.3	+34.7		
47	All	−1.7	−11.7	40	54	56,57	All	+52.6	+12.6	75	40
	1	−57.2	−14.2				1	+2.2	+20.1		
	2	+24.5	+6.2				2	−28.4	+44.2		
48	All	−18.9	+3.2	80	33	58	All	−20.7	+3.2	43	67
	1	−65.7	+23.3				1	+9.4	−45.7		
	2	−5.4	−18.8				2	−40.9	+27.1		
49	All	−2.0	+8.4	55	50	59	All	−3.5	+0.7	64	50
	1	−47.3	−31.4				1	+12.1	+9.1		
	2	+63.2	+13.8				2	+9.5	−41.3		
50	All	+99.4	+8.2	36	56	60,13	All	+104.8	−18.0	50	50
	1	+32.0	+4.5				1	+105.8	−19.9		
	2	−81.1	−8.6				2	+157.0	+13.0		
61	All	−19.2	−13.3	100	50	71	All	+9.1	+5.9	51	42
	1	−33.6	−23.2				1	−3.4	−0.3		
	2	−15.4	+17.3				2	−24.0	−33.5		
62	All	+19.4	−11.6	67	71	72	All	−50.6	−3.8	100	100
	1	−32.4	−30.2				1	−54.8	−11.8		
	2	+10.7	−17.2				2	−21.5	11.5		
63	All	+17.3	+8.2	25	0	73	All	+27.3	+2.0	72	65
	1	−42.1	−39.6				1	−8.1	−9.0		
	2	−21.9	−0.2				2	+0.5	−15.3		
64	All	+15.8	−0.4	33	50	75	All	−25.2	−6.9	80	89
	1	−25.5	+44.6				1	−42.5	−7.2		
	2	+9.8	+4.6				2	−60.3	−36.4		
65	All	−16.5	−4.4	73	76	76	All	+46.6	−14.9	100	67
	1	−22.7	−23.1				1	+6.9	−18.7		
	2	+22.7	+27.8				2	+278.4	+33.7		
66	All	+43.8	+16.7	68	38	77	All	−18.9	−4.1	0	67
	1	+39.3	+1.8				1	+26.4	−28.7		
	2	+113.2	+30.9				2	−34.6	−2.3		
68	All	+55.1	+9.6	67	76	81	All	−12.8	−10.8	41	81
	1	+34.9	+6.7				1	−42.7	−40.8		
	2	+107.9	−12.0				2	−31.4	−7.4		
69	All	+16.3	−6.5	72	71						
	1	+31.7	−4.7								
	2	−39.9	−9.8								
70	All	−4.5	−18.1	72	75						
	1	−4.2	−15.4								
	2	−15.9	−11.5								

[a] Percent change in the primary direct coefficients for the largest customer of the given producing industry.
[b] Percent change in the primary direct coefficients for the second largest customer of the given producing industry.

Column 4 presents the same information for the period 1958–1963. Columns 5 and 6 give the percent of the consuming industries which showed coefficient changes in the same direction as the change in the weighted average of the coefficients for all the customers of the given producing industry during each period.

An examination of the data in Table 5 on individual input coefficients indicates a very marked degree of variability in the extent and direction of change in both subperiods. Industries that showed positive changes in overall intermediate requirements for their output during both 1947–1958 and 1958–1963 often showed frequent decreases in coefficients into individual consuming industries. Similarly, industries experiencing overall negative changes in direct requirements for their output showed, in both subperiods, numerous instances of increasing requirements for particular industrial consumers. This opposite direction of movement was not necessarily confined to the smaller customers of a given industry's output. Columns 3 and 4 of this table indicate considerable evidence of individual coefficient change in the two largest intermediate customers of an industry's output which was opposite that of the overall average.

While there appeared to be some tendency for the majority of the customers of a given industry to show direct coefficient changes in the same direction as the overall change in both subperiods, this tendency was not marked. Between 1947 and 1958 of the 70 industries in the analysis, only 44 showed coefficient changes of the same direction as the average for the majority of their consuming industries, while between 1958 and 1963 for only 51 of the 76 industries examined did the coefficient changes for the majority of the consuming industries adhere to the overall change. For both time periods, there was little conformity of individual coefficient changes to the average change, even in those industries which showed large average changes over the period in the direct requirements for their output, nor were the causes of opposite direction of movement confined to those industries with small average changes.

VI. CONCLUSIONS

Here we examined interindustry structural and coefficient changes in the American economy between 1947 and 1966. We found that the total domestic output required to produce 1958 final demand with different interindustry structures between 1947 and 1966 differed little. However, the average change in output requirements from individual industries varied widely, with those from many industries increasing faster than the average and those from other industries decreasing more slowly than the average, or actually decreasing.

The average change in total intermediate output requirements (as measured by inverse coefficients) was larger in each subperiod than the average change in total output requirements, but the rate of change of intermediate output requirements was highly erratic between subperiods. There is little evidence of a slowdown in the rate of change of intermediate output requirements after 1958, as had been previously hypothesized. Even considering only the benchmark data, the average annual rates of change in intermediate output requirements experienced by most industries between 1947 and 1958 and between 1958 and 1963 were highly erratic and uneven. Not only did the magnitude of the average annual rates of change for most industries vary widely, but for one-third of the industries even the direction of the rate of output change differed between 1947–1958 and 1958–1963. Further, between 1947 and 1966 the relative impact of structural change on the output requirements of different industries was similar irrespective of whether it was measured as the impact on total output requirements, as the impact on intermediate output requirements derived using inverse coefficients, or as the impact on intermediate output requirements derived using primary direct coefficients.

Between 1947–1958 and 1958–1963 the cell-by-cell input coefficients for the various direct customers of most producing industries did not change in the same manner as did the overall average. There were numerous cases where either the majority of input coefficients or the input coefficients for the largest customers of an industry (or both) did not move in the same direction as the overall average for the industry.

These findings have some important implications for empirical work based on US input–output data. First of all, the major conclusion which emerges here is that postwar interin-

dustry structural and coefficient change in the American economy has been highly erratic, uneven and unpatterned. This implies that the accurate forecasting of structural changes or of changes in individual input–output coefficients is extremely difficult. For most industries few clearcut trends or patterns of change in output requirements were evident over the two decades analyzed, and projections of changes in industry output requirements on the basis of historical rates of change will thus not be likely to produce acceptable results.[10] Further, as discussed in Section V, in most cases the average pattern of change for most consuming industries was not characteristic of all or even the major customers of that industry. Thus, attempts to adjust individual input coefficients on the basis of movements in industry outputs or value added, such as the RAS method, are not likely to produce accurate coefficient estimates.[11]

Another major implication of these findings is that, while the input–output coefficients for some industries tend to remain stable over time, this is usually the exception rather than the rule. The input coefficients for some industries may tend to be stable over certain periods but change markedly over others. This indicates that the interindustry structure of the American economy for a certain year should be used to approximate that of another year only with a high degree of caution. This is expecially true as the time period between the two years in question lengthens.

Thus our findings here are largely negative, and it is clear that much additional research will be required before we will be able to model structural and coefficient change over time and predict these into the future. One useful area of future research would be to determine how much of this seemingly erratic change is due to identifiable factors such as price induced substitution among inputs, changes in product mix, and changes in the technological requirements of production processes.[12] Also, given the lack of a historical times series of observations on US input–output coefficients, it would be useful to relate changes in industry output requirements and in individual coefficients to movements in those GNP components for which we do possess reliable time series data.[13]

Before concluding, however, it is necessary to consider whether even ambitious research efforts of the types recommended here would be sufficient to make input–output an accurate and reliable forecasting device. While the input-output model does allow us to disaggregate the economy into detailed industries, it requires some very rigid assumptions such as product homogeneity and the proportionality of inputs to outputs. These assumptions themselves may be largely responsible for many of the seemingly inexplicable findings reported here. In essence, it may be unrealistic to assume that the economic–technological structure of the economy during any given period of time will be an accurate representation of that structure several years hence—even if the technical coefficients have been modified or projected. The production functions of the real world may differ considerably from the linear Cobb–Douglas production functions implicitly assumed in input–output theory. In any case, it is important that the potential user of actual or projected input–output data be aware that, theoretically, the basic assumptions of the input–output model may be questioned, and that empirically, there is much about interindustry structural and coefficient change which we do not understand.

[10] One such method utilizing indexes of movement computed on the basis of the average annual rate of change of intermediate output requirements is developed in Shapiro [9].

[11] The RAS method of input coefficient estimation is developed in Stone [10].

[12] The results of a recent study [2] indicate that shifts in product mix may not be as important a cause of changes in interindustry relationships as is generally believed.

[13] Attempts have already been made to impute industry detail into aggregate macroeconometric models, such as in the Wharton model [8] and the Brookings model [7].

ACKNOWLEDGEMENT

The author is grateful to Beatrice Vaccara, Jean Frazier, Stephen Winkeljohn, Philip Ritz and Robert Wendling for advice and assistance in the course of this study and to a referee for comments on an earlier draft of this paper but retains sole responsibility for the opinions expressed here and for any errors. This work was supported, in part, by the Federal Government's Interagency Economic Growth Project, and was conducted while the author was with the Bureau of Economic Analysis of the US Department of Commerce.

REFERENCES

1. Bezdek RH, Frazier J & Wendling RM (1974) Documentation for the deflation of the 1966 input–

output matrix to 1958 dollars. US Department of Commerce, Bureau of Economic Analysis (unpublished).
2. BEZDEK RH & DUNHAM CR (1976) Changes in industry product mix as a cause of intertemporal instability in input–output coefficients. *Q. Rev. Econ. Bus.* **16**(1), 61–76.
3. CARTER AP (1970) *Structural Change in the American Economy*. Harvard University Press, Cambridge, Massachusetts.
4. FRAZIER JR (1974) Methodology used to reprice the 1963 input–output domestic base tables in 1958 prices. US Department of Commerce, Bureau of Economic Analysis (unpublished).
5. GOLDMAN MR, MARIMONT ML & VACCARA BN (1964) The interindustry structure of the United States: A report of the 1958 input–output study. *Surv. Curr. Bus.* **44**, 10–17.
6. KLEIN LR (1974) Econometrics of inflation: 1965–1974. Paper presented at the NBER Conference on Research in Income and Wealth, Bethesda, Maryland.
7. KRESGE DT (1969) Price and Output Conversion: A Modified Approach, in *The Brookings Model: Some Further Results* (Eds DUESENBERRY JS *et al.* Rand McNally, Chicago and Amsterdam.
8. PRESTON RS (1972) *The Wharton Annual and Industry Forecasting Model*. Studies in Quantitative Economics No. 7, Wharton School, University of Pennsylvania, Philadelphia.
9. SHAPIRO AK (1972) *Input–output analysis as a predictive tool*. BEA Paper No. 20.
10. STONE R (1962) A long-term growth model for the British economy. In *Europe's Future in Figures*, Ch. 10 (Ed. GEARY RC). North-Holland, Amsterdam.
11. US Department of Commerce, Bureau of Economic Analysis (1969) Input–output structure of the US economy: 1963. *Surv. Curr. Bus.* **49**, 16–47.
12. US Department of Commerce, Bureau of Economic Analysis (1970) The input–output structure of the United States economy: 1947 (unpublished).
13. US Department of Commerce, Bureau of Economic Analysis (1972) Input–output transactions: 1966. Staff Paper in Economics and Statistics, No. **19.**
14. VACCARA BN (1969) Changes over time in input–output coefficients for the United States. In *Applications of Input–output Analysis* (Eds CARTER AP and BRODY A) pp. 238–260. North-Holland, Amsterdam.
15. US Department of Commerce, Bureau of Economic Analysis (1969) Changes over time in US input–output relationships. Paper presented at the Seminar on Input–Output Analysis sponsored by the YMCA Center for International Management Studies in the State Planning Committee of Romania, Bucharest.

ADDRESS FOR CORRESPONDENCE: *Dr. Roger H Bezdek, Division of Solar Application, Department of Energy, Washington, D.C.* 20545, *USA.*

ol. 6
978

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 227–230
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0227$02.00/0

The Use of Formal Inventory Control Models: A Preliminary Survey

AP MUHLEMANN

University of Manchester Institute of Science and Technology, UK

AG LOCKETT

Manchester Business School, UK

(*Received September 1977; in revised form November 1977*)

Of the many papers and texts on the subject of inventory control, relatively few report the successful implementation of systems based on formal mathematical models developed with specific objectives in mind. This could be considered rather strange in the light of the relatively high number of papers reporting the development of models for dealing with a multitude of different hypothetical situations, and the widespread availability of computer systems capable of exploiting such models. Various reasons for this anomaly can be put forward. This paper reports the results of a preliminary survey designed to investigate this more deeply, and draws some necessarily tentative conclusions relating to factors contributing to the successful application of such systems.

INTRODUCTION

INVENTORY control is possibly one of the most widely written about topics in the management science/production management areas. There are numerous texts which cover various aspects of the topic (see, for example, Hadley and Whitin [6], Lewis [8]), and it is unusual to pick up a journal on management science which does not include at least one article which could be classified under the broad heading of Inventory Control.

However, of these books and papers, relatively few seem to report the successful implementation of systems based on formal mathematical models developed with a specific objective in mind. That is not to say that this is the only approach leading to an efficient system, but it could be considered rather strange in the light of the relatively high number of papers reporting the development of models for dealing with a multitude of different hypothetical situations and the widespread availability of computer systems capable of exploiting such models. (For a comprehensive survey of inventory theory, and its applications, see Aggarwal [1].)

Many reasons for this anomaly can be put forward, ranging from the simple proposition that organisations with successful systems did not want to broadcast the fact, in order to stay ahead of competitors; to the proposition that mathematical models can never cope successfully with all the complexities and uncertainties of the real world. There are clearly some factors which, taken at face value, should contribute to the successful implementation of sophisticated inventory control systems. These include a lack of uncertainty or at least unquantifiable uncertainty in an organisation's operations, availability of staff who can develop and sell such systems, adequate computing facilities to exploit such systems.

In this paper we report the results of a preliminary survey looking at the inventory control systems used by a number of organisations, trying to measure the extent to which formal inventory models are being used, and the factors influencing successful implementation of such models.

THE SURVEY

A number of other authors have carried out surveys in areas related to that discussed here. These range from the more general type of survey into the application of quantitative techniques in general (included in Eilon *et al.* [3]

and Davis [2] for an American survey); to those which concentrate on a specific topic (see Ellis *et al.* [5], who survey the applications of Dynamic Programming, and New [9] who looks at the Management of Manufacturing Operations).

The majority of these and other surveys have relied upon using postal questionnaires to obtain the information required for their analysis. Postal questionnaires have a notoriously low response rate, some authors quoting as low as 30% usable replies; with a carefully chosen sample, this can be quite a disruptive influence. In order to attempt to overcome this, and to reduce the time scale of the project, the information was primarily obtained by personal interviews, based on an information checklist. This also avoided another problem with postal questionnaires: involved questions requiring complex replies obtain poor response, dichotomous questions which are simple to answer can conceal the truth. The interview itself was arranged by telephone, and carried out by two mature students in the Department of Management Science at UMIST, both of whom had practical experience of inventory control. Only one company refused to see our interviewers.

A number of factors were decided on to limit the size of the sample. The firms chosen were in the Manchester area, owned or rented a computer, and were listed in Kompass [7], and the sample of 25 companies covered a range of sizes and industries. Some of the findings of the survey are now presented below, without detailing any individual company.

A SUMMARY OF THE RESULTS

The twenty five companies were divided into four broad product groups:

i Engineering (e)
ii Chemical and Allied (c)
iii Textiles and Clothing (t)
iv Building Materials (b)

Each company was then classified as small, medium or large according to number of employees. This resulting breakdown is given in Table 1.

As a very crude measure of effectiveness of inventory control, we looked at the average annual inventory value, expressed as a percentage of annual turnover. These results are presented in Fig. 1.

TABLE 1. BREAKDOWN OF SAMPLE BY PRODUCT GROUP AND SIZE

Group	Number of companies of size: small	medium	large	Total(s)
i	1	6	4	11
ii	2	1	2	5
iii	5	1	1	7
iv	1	0	1	2
Total(s)	9	8	8	25

Key: small: 749 or less employees; medium: 750 to 2999 employees; large: 3000 or more employees.

No significant picture emerges from this, with the exception of an indication that possibly smaller companies (with smaller inventories?) control their inventories more efficiently. However it might be that the measure considered here is too crude.

Next we moved on to look at the level of sophistication of the inventory control system operated by companies. We classified the systems on a five point scale, as follows:—

1. Manual system, but no mathematical basis
2. Manual system, but with mathematical basis
3. Computerised stock records, but reorder calculations carried out manually
4. Fully computerised system using a formal inventory model and crude forecasting techniques

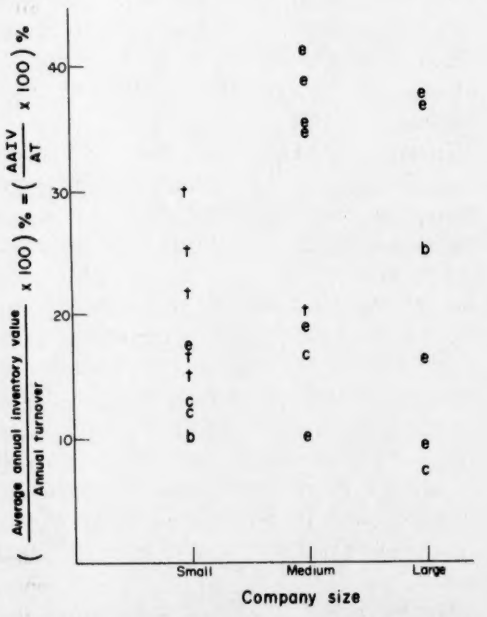


FIG. 1. *A measure of inventory control efficiency against company group and size.*

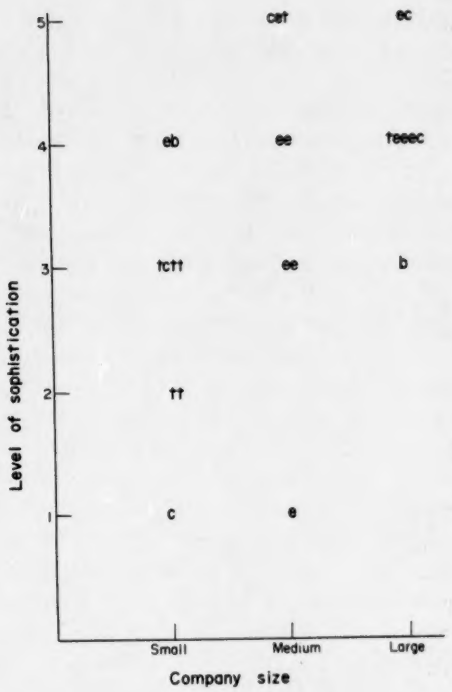


FIG. 2. *Level of sophistication against company size.*

5. Fully computerised system using a formal inventory model and sophisticated forecasting techniques

Figure 2 shows the level of sophistication against company size.

This indicates that overall, larger companies tend to have more sophisticated systems. There seems to be no evidence to suggest that one group is significantly in front of or behind the others in terms of sophistication.

Figure 3 shows inventory control efficiently (measured as in Fig. 1) against level of sophistication. There is slight evidence that overall more sophisticated systems produce an improvement in this measure of efficiency.

COMMENTS

The classification of systems used in the previous section is simply based on the information the individual companies were prepared to release, and does not take into account the appropriateness of the models being used in practice. Although companies whose systems were classified as category 4 or 5 were apparently using quite sophisticated systems, this does not necessarily imply that the models being used were appropriate for that environment. Moreover several of the companies were unwilling to discuss details of the models, and in these cases the classification tends to be more subjective, and on the generous side.

The majority of all the systems were either of the max/min or the reorder level, reorder quantity variety. Of interest to us is the actual basis of the decision over whether or not to place an order, and the basis of the calculation of the quantity to be ordered. The 'favourite' way of expressing these parameters was in terms of forecast demand, and the following are some typical examples from the category 4 classification.

i. reorder level = 3 months forecast demand
   reorder quantity = 2 months forecast demand
ii. maximum level = 20 days forecast sales
   minimum level = 4 days forecast sales

Clearly of paramount importance to these sort of systems is the method by which demand/sales are forecast (see [4] and [10] for a discussion of this point). In the study the methods ranged from sophisticated forms of exponential smoothing to crude, manually inputted estimates. Even after the 'selection' of a satisfactory forecasting method there is still the question of how many months/weeks/days

FIG. 3. *A measure of stock control efficiency against level of sophistication of inventory system.*

forecast demand should be used for the reorder level/reorder quantity. Often, when companies were willing to divulge the basis for this, it transpired that they made the decision on the basis of subjective judgement rather than economic criteria.

Some other companies had developed quite complex formulae for reorder quantities, but were apparently unable to explain their origins. One company was using the 'classical' EOQ formula with a stockholding cost calculated on the basis of 12% of the value of an item, at a time of very high inflation.

It was not the intention to suggest that EOQ type formulae should be used, but to find out what is used. Although a number of formal and apparently very useful systems have been developed and implemented, the reasoning behind the criteria for reordering and the quantity to be reordered seem often to be rather arbitrary and based on 'company tradition' rather than economic consideration. Moreover, in a number of cases the users considered themselves to be using an economic order quantity because it was based on such considerations as a head office directive that average stock levels should not exceed two months demand. This being based on a strict financial limit obtained at a broad level where the detailed implications of higher or lower limits had not been considered.

CONCLUSIONS

The most conclusive result was that larger companies tended, naturally enough, to have more sophisticated systems. Whether more sophisticated systems result in 'better' control of inventory is another matter. There is a real difficulty in measuring the 'goodness' of inventory systems against each other, designers are prone to claim multitudes of intangible unquantifiable benefits. Moreover, within the same product group one organisation may be dealing with a deterministic environment while another might be faced with a complex and uncertain environment. Thus there is the problem of finding a common bench mark against which we can honestly measure inventory holding efficiency.

There did appear to be some evidence that some organisations were using quite sophisticated formal models for inventory control, and especially the forecasting aspects of the systems. These tended to be the larger organisations where 'scientific management' had already gained a substantial foothold, and hence management had been educated to accept this approach. In those smaller organisations with sophisticated systems, there tended to be a person in a key position responsible for the system, with the personality, drive, enthusiasm, ability to convince others to accept the system, and work with it. Moreover there was evidence that those companies with sophisticated systems felt they were ahead of competitors, and were therefore reluctant to discuss their models in any real depth.

Finally it appears that the majority of companies are deriving the main benefit from accurate stock recording and sales forecasting using computer systems, rather than formal mathematical models based on economic objectives. There is little evidence to show much use of such models—or even desire from the companies used in this survey. It may be that a much larger survey would give different results, but we doubt if the main conclusions would significantly alter.

Vol. 6
1978

REFERENCES

1. Aggarwal SC (1974) A review of current inventory theory and its applications. *Int. J. Prod. Res.* **12**(4), 443–482.
2. Davis E (1974) APICS state of the art survey, production and inventory management. **15**(4), 1–11.
3. Eilon S, Hough J. & Betts RJ (1969) Profile of current members of the operational research society. *Opl. Res. Q.* **20**(2), 247–273.
4. Ekawayake RB, Norman JM, Richie E, Smith PH, Taylor BR & Wilcox P (1977) Aspects of an inventory control study. *Europ. J. Opl Res.* **1**(4), 225–229.
5. Ellis C, Lethbridge D & Ulpa A (1974) The application of dynamic programming in United Kingdom companies. *Omega* **2**(4), 533–541.
6. Hadley G & Whitin TM (1963) Analysis of inventory systems. Prentice-Hall, New Jersey.
7. Kompass (1976) *Register of British Industry and Commerce*, 14th Edn, Vol. 2. Kompass Publishing, Croydon.
8. Lewis CD (1972) *Scientific Inventory Control.* Butterworths, London.
9. New CC (1976) *Managing Manufacturing Operations.* British Institute of Management, London.
10. Richie E & Wilcox P (1977) Renewal theory forecasting for stock control. *Europ. J. Opl Res.* **1**(4), 225–229.

ADDRESS FOR CORRESPONDENCE: AG Lockett Esq, Dean, Faculty of Business Administration, Manchester Business School, Booth Street West, Manchester M15 6PB, UK.

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 231-236
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0231$02.00/0

The Role and Needs of the Management Science Practitioner

PADDY FITZGERALD

(Received February 1978)

MOST management services practitioners are aware of the formidable list of deficiencies levelled at them time after time:

- Management Science covers the inessential problems;
- Its work is based on unrealistic assumptions;
- Science cannot include the necessary human and social factors;
- Specialists cannot communicate with managers;
- Managers feel threatened by Management Science.

Many will probably nod their heads with greater or lesser emphasis according to the environment which has suffered their labours—for managers and working situations are at least as diverse as Management analysts.

If, on the other hand, one is tempted to rise up to defend the profession, for so it is, against the specific complaint—yet one must agree, perhaps reluctantly, that deep in one's heart there is a sense of disappointment : the disappointment one has towards a gifted pupil whose abilities seem evident and yet who somehow fails to deliver the goods.

Of course, if one refers to the tools now in common use and become particular, asking 'is there disappointment over the results from networks, from mathematical programming, from simulation—or from any other of the techniques so familar to many of the educational bodies in the business?', then one would have to answer 'No—quite the reverse' (which is probably why they are still taught).

So wherein lies the disappointment?

Well, let me look at things another way....

The Manchester Business School in Britain (and perhaps others) had been using Rolls Royce as a case study of incipient failure for two years before the failure occurred. Now why, for longer than that, given their inside information, had the internal management scientists not been beating—in Cassandra fashion—at Management's door? Had they, and were they just not heard above the roar of slide rules?

Or again, one might ask how many management scientists contributed to Lord Ryder's report on British Leyland, or to the House of Common's Expenditure Committee's report for that matter. If, gentle reader, your lips have curled, no doubt at the thought of the standard of analysis displayed in those reports, then our omission is doubly damning.

Having mentioned British Leyland (or BL for short), let me pursue this Company a little further. There was a time when it seemed that any self-respecting newspaper offered us daily a new solution for the patient's illness (that the patient was ill was about the only thing people agreed upon). Solutions advanced included:

- It was too big : just beyond the abilities in size for any one Board—one ought to split it up. Yet there are plenty of automotive companies covering as wide a market or markets with turnovers bigger than theirs. Whether one considers numbers employed, capital employed, number of sites managed or any other criteria of size, one will still

be able to find other companies bigger and more successful.

- It has 'difficult' workers—but for instance the Bromsgrove area where I live supplies similar numbers to Longbridge and to Garringtons (where I used to work and where I know labour problems have been slight): are Leyland supremely capable only in their bizarre ability to pick trouble makers?

- Its management is poor. But if one reviews their structure carefully one finds a good sprinkling of the best Ford could produce (and they are held up as the standard—the invidious comparison) as well as former employees from Massey Ferguson, Sperry Rand and other companies respected for their management ability.

- The marketing is bad. An argument advanced powerfully by Michael Shanks in March 1977 in the *Observer*, in an article which ranged from investment to sales channels. Well, investment is no doubt a problem, but Leyland's problems over the last years have not been fundamentally in this area : try today to get a Land Rover, Range Rover, Rover 3500, 5 speed TR7 at all, or Daimlers, Jaguars or Allegros in *your* choice of colour and trim.

Without labouring the point further, let me put forward an alternative analysis of their problem, which I believe, is firmly rooted short term in the production area. I can do this best by describing two work situations:

1. Overall Control: At a fixed point in time, marketing provide forecast quantities of sales for the coming year. Production use these to deduce a budget *total* cost (excluding raw material) for the product mix envisaged.

This cost is used both by sales for pricing and by the Company as the control on production, which is a *budget total cost* versus *actual total cost* for the location.

If demand drops off and sales cry off their numbers, the pressure is off production (until the next budget) and one can see that targets are likely to be met by this function. On the other hand if demand has been *underestimated*, can you see a production director upping volumes and getting a hiding for it?

If you doubt me try asking any plant director what the unit cost was over the last time period for any of the variants he produces. He will not be able to answer you.

2. Examine the power train production system for Longbridge (as an example again) and you will find that in the best fashion this is based on parts explosions of the expected product mix. Unfortunately, two outside bodies, BL International (the export side) and Parts and Service will requisition at short notice and rob the line.

Since, apart from arrival checks (really a check on suppliers), there is no further check on material— the first opportunity to discover the shortages is when the man screwing dimple holders into engines runs out of dimple holders—at which stage the chain is empty right up to him. What I am saying is that no-one knows exactly how much of anything is in the pipeline—which, with their extensive conveyor system is a large quantity. Of course, BL has built up remarkable relationships with their suppliers and, hopefully, dimple holders are soon on their way: but if you were on the line, and particularly if your pay depended on it, how would you feel about this? In fact, many car workers are aware that, although their pay rates are higher than average, their average pay over a year is not, and this difference which is often outside their control is clearly a major frustration.

What I am suggesting then is that Leyland's problen is simply a gross inadequacy of information, systems and feedback controls. The new Ford talent, bereft of the backup to which they are accustomed, have their teeth drawn—the problem is simply too vast for any of them individually to tackle; indeed, it is so big that the individual down the line does not recognize it.

Of course it has not always been so. Each of BL's inherited companies had some sort of systems of their own, capable then of coping; but the problems of running all the units as one business, and to run them as has become fashionable, in a much more integrated manner, poses problems which are beyond current capabilities. Leyland's problems in information and control outstrip the problem of re-equipping which has received so much priority and attention. It seems obvious, does it not—a

Vol. 6
1978

tighter operation made possible by better control—one cannot be had without the other.

If things are *so* simple, why have they not been recognized?

Well, that pontiff of management, Peter Drucker, in his book "Management Tasks, Responsibilities and Practices" puts it this way:

> "....Economic progress can be defined as the ability to take greater risks. The attempt to eliminate risks, even the attempts to minimize them, can only make them irrational and unbearable. It can only result in that greatest risk of all—rigidity."

In other words: management has to do with risk taking. Risk taking has to do with uncertainty. Therefore, the higher the management, the *more uncertainty there ought to be.* Typical of this is the man who says he 'runs the place by the seat of his pants'. This is an expression borrowed from flying to distinguish the natural pilot fron the plodder. It is a fortuitous analogy, since today pilots are aware that, while this might be fine in clear conditions the best of them has to rely on instruments in cloud, or that is, when the area he is moving towards is unclear.

Let us change analogies not to strain them, moving from one element to another with ease. Consider steering a ship. The bigger it is, the longer it takes to turn (as the Captain of the Torrey Canyon discovered) and the longer one has to anticipate the need to turn. But venture on to the bridge of a mighty ship: you will find a helmsman who is *constantly* moving the wheel. The ship veers gently one way then another in answer to random forces and, eventually, to the helmsman's movements of three minutes ago; that doesn't stop his behaviour and he is happy because he feels the changes he is making are having immediate effect. Well, management can be like that—because it is travelling in fairly thick fog—it often fails to appreciate the size and inertia of what it is trying to steer, but it is happy moving the wheel just a bit—and frequently as well, so that everyone realises there is someone up there! And as the manager has moved up, so he has become inured, one might say anaesthetised, to poorer and poorer relevant information—he has come to regard his situation as a fact of life. Judgement is no substitute for analysis. After all, judgement should be more than informed guesswork, and the two should not be confused.

Again parenthetically, one can look at British Airways as a contrast. We will come back to them in a wider sense, but it is clear that the integration between British European Airways and British Overseas Airways Corporation was greatly eased by the fundamental need of each separately and both together to have an accurate and reliable information base as an essential to routine operations—in this climate one can start to produce management tools that are not built on sand—tools that management can believe in and wish to develop: and this accounts for what might appear to be the enlightened attitude in British Airways management.

Or, to choose another example, reading the annual report of British Aluminium one sees that the organisation is going through what seems a very positive phase (profits leaping upwards), but one also sees that organisationally they have decentralised between 1969 and 1973. Now that centralisation may well have been bad in human terms, but perhaps their present fortunes are built on effective information and systems created in this era. Obviously, one *can* have corporate systems with autonomous units—but it is that much more difficult to achieve.

One last example concerns my old Company Guest, Keen and Nettlefolds (GKN). At one stage a while ago they were concerned (as were most companies) with the question of cash flow. They wanted the maximum effect with the minimum of interference: in fact, simple analysis (undertaken rather too late I'm afraid) showed that the three major spends to each Company were the steel bill, the energy bills and the inter-company transfers—all three dictated by the centre. Manipulating the dates of one or more of these would have been enough to change the behaviour of individual companies to give a Group cash flow picture that would certainly have been better than the event. For, devoid of proper information (despite computerisation of many areas the individual accounts were handled on individual accounting machines—a method which allows one to see each tree with great clarity), the Group was forced to rely on the broad blunt rule: reduce working capital (stocks) by a fixed percentage. This had a number of effects:-

a. Those companies which were the most effi-

cient suffered most, for it destroyed the balance between stock and bulk purchasing;

b. The easiest way of reducing stocks is to stop ordering raw materials, which

(i) creates havoc downstream in a vertically integrated Group;
(ii) reduces one's delivery performance, usually in the most popular products at a time when delivery is a prime ingredient for success in the market.

Incidentally, in this example a determined assault on obsolete stock may well have been much more successful in generating cash, even if it would also have required paper losses in the accounts.

This presents one of the fundamental limitations of Management Science—it can usually be ignored without chaos ensuing: no-one is held to account for what might have been, and if you want an easy life, doing a bit better than last time is more than enough.

Belatedly, we are ready to state our first crucial need, namely a new awareness in much of our management of the need for facts and data. It is a prerequisite to a management sense of purpose, because then one can home in quickly and effectively on the issues of importance. This lack of proper routine data is, I am sure, the major reason for the rift between practice and academic teaching.

The second crucial need, is for management science to be represented at the right level. Inevitably management science takes time to product results. Someone has to take the pulse, judge the timing, deduce the relevant and present the case in terms of the arguments that carry—at that time—the greatest weight and, most important of all, someone has to anticipate. Notice that the attributes I have given most weight to are those which are rarely taught or addressed in a management course. That does not mean that they are not capable of being taught, merely that few teaching institutions are making the attempt. Let us list these attributes:

- Sense of timing and relevance
- Anticipation
- Perspective
- Insight
- Judgement, in order to engender mutual respect
- Entrepreneurial ability

Returning to the main argument, we have already noted that scientists are often accused of solving the wrong problems. My experience is that this is usually because someone else has already decided what the problem is—most consultants today operate on the premise of 'getting into' an organisation: analysis and persuasion come very much second to being allowed to contribute. If I may take a personal example—GKN has Swedish companies and during my time with them I eventually engineered an invitation from the Financial Director to visit and advise on the most effective way of using working capital tied up in stocks between raw material, work in progress and finished goods. The chief executive said at our first meeting "You are here for a week—go where you like, ask what you like and don't feel limited to the stock problem. Come and see me before you leave".

Opening up the problem led to a much more fundamental problem which until then had not been recognised. This was that the lack of adequate data meant that the systems created over time were crude and in fact were leading them to be very competitive for small throughput items and much less competitive for the large volume items that the factory was really suited to produce. Over the years the product mix had been moving in quite the wrong way for the strengths they really offered to the market.

Hence the inclusion of the entrepreneurial element in my list of attributes....to add the problems that have not been spotted to those that have. Rolfe Tomlinson, while Managing Director of the National Coal Board's Operational Research (OR) Executive, carried out a survey which showed that the Board felt that the most valuable OR work his Unit had done was often the most difficult to quantify in terms of benefits. I think this brings two of my strands together: it is difficult to quantify because the data did not exist before the work began and one could not, therefore, make a before and after comparison which was more than subjective—and, secondly, because the

data was missing, line personnel had missed the problem....probably over quite a long time period.

In a recent case study, we went back to the Managing Director of GKN Lincoln Electric to find out what results he felt had now been achieved. The numbers were impressive:- a 27.3% reduction in stock, equivalent to £351,000 in 1975 prices. But in allocating benefit he said:

1/3rd was saved by the OR work directly;

1/3rd was saved on the increased level of information peripherally produced by the system;

1/3rd would probably have been saved later if not sooner due to constant reviewing of factory operating procedures.

Thus, the two things a practitioner is searching for are significant savings and a major impact on operations—and these represent opportunities which need to be ferreted out and recognised, hence the entrepreneurial approach.

An unexpected bonus from getting the level right is that more of a career path is open to the specialist. One of the profession's problems is that although people flock to get into it, they also flock to get out. I am not alone in having shed the words Management Science from my title—yet the skills and training I am deploying are basically the same: indeed, if you removed all OR men from the planning field it would become a bare field indeed.

Many managements realize that a large project of capital expenditure puts unaccustomed strains on routine operations and, though financial control is often strict, the questions of getting new investment performing to the right time scale is often neglected. An involvement with the approval *and implementation* process for investment would give the planner practical experience which would be seen as such by line personnel.

So what I am advocating is a function which would be the focus for change—short or long term, and this position, reporting at the highest level, would bring as much benefit to the firm as the appointment of a finance director instead of just an accounts manager.

And so to the third crucial need, that of management. We are already at a point where the simple criterion of profits has been overtaken and where the speed at which conditions and attitudes change demand new batteries of aids for decision making.

Let me cite three examples: The first concerns British Airways who carried last year roughly as many passengers as the year before over roughly the same number of passenger miles. But in many cases, due to the system passengers had no real choice in carrier or in price. For most business traffic at least, the demands of timing dictated whose aircraft would be used. On the other hand, from the accounts one can see that the significant improvement in surplus (on which presumably the Government will judge the Company) has been through reducing operating expenditure, not by increasing the load factor. Clearly purely financial criteria would lead us to be subjected to poorer and poorer levels of service.

In the same industry, the pressures of finance are allowing the British Airports Authority rapidly to become a significant supermarket chain...if nothing else. Once again the Authority may be the financial standard for Europe, but ask any traveller what he thinks of the process of arrival or departure at Heathrow Airport compared to other major European airports. It is certainly no defence to say that Heathrow handles more traffic.

My next example is of an article in the Financial Times on 5th September, 1977 by Colin Jones called "Setting profit targets in State Industries". He states '....the emphasis is upon the wrong objective—upon the maximisation of profit rather than upon the minimisation of cost.'

Most of us could produce many examples where the cost minimisation attitude is plain foolish and in many cases has led to public hostility. Almost every public service today, in its attempts to cut down on costs are trimming the use of capital intensive services while maintaining or increasing the administration levels in true Parkinson manner.

My third example is from investment. British companies have been more successful in general than Continentals at squeezing higher rates of return from their projects. But cursory knowledge of DCF techniques shows that the higher the approval or filter rate, the shorter the time scale for the project.

One might speculate on whether this is one of the problems today—whether anyone can

justify the long term investment any more, on which depend the myriad of short term juicier proposals, whether in fact, perhaps for other reasons, the investment scene is beginning to resemble that for developing countries. Couple this with the poor levels of utilisation and the erosion of the UK manufacturing base comes into focus.

It is time to gather the strands of the argument and to look ahead. We need to produce a new awareness of our ability to contribute and we have to increase the level at which we report. Also there is in many companies an urgent need for more and better information and we have a part to play in this as well. I do not believe that this is special pleading. OR has dealt with many of the lower level problems. The problems of the future are inevitably problems at a higher level—as we come to try and squeeze more and more from what we have.

To critics who claim that our contribution is and will be limited because we do not include in our models human and social behaviour I will say just this: modelling of human behaviour will come, but there are many cases where the human understanding is needed not within the model, but in appreciating the art of the possible to begin with, and in developing the necessary sensitivity when it comes to getting the work inplemented. The human skills of the practitioner are much more necessary at this point in the development of management science than is the capability to model human factors.

The bulk of our work will be first to provide tools that are relevant and are of immediate use, operating within as short a time scale as possible to allow meaningful examination of options and to allow managements to explore the assumptions themselves: our job is to clarify and illuminate the decision making process, not to take it over. Secondly, we have to do work which crosses boundaries, shifts between functions, shifts between locations and between business environments according to the problem.

In this sense real time computing has given us a chance comparable to the discovery of North Sea Oil. We have the ultimate participation and communication tool: we will either squander this opportunity and be relegated for good to the company backwaters, or we will be in the very van of progress. We need to make a change which is as big as the one which faced management with the introduction of the computer but we need to achieve more than the equivalent of computerising the payroll program.

ADDRESS FOR CORRESPONDENCE: *Paddy FitzGerald Esq, The Old Vicarage, Cofton Church Lane, Barnt Green, Worcestershire B45 8PT, UK.*

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 237–247
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0237$02.00/0

Cybernetic Approach to Investment Decision Making

J FELSEN*

St John's University, Jamaica, NY

(*Received March* 1977; *in revised form July* 1977)

This paper summarizes the results of our research into applications of cybernetic concepts and artificial intelligence techniques to investment analysis: it outlines the philosophy underlying the cybernetic approach to market forecasting and investment selection. This approach is computer oriented—it enables us to automate or program directly the complex *judgmental* aspects of investment decision making. The cybernetic approach alleviates some deficiencies of conventional statistical methods. Specifically, it (1) explicitly includes the time dimension into investment analysis, (2) is based on methods for decision making under *uncertainty* rather than risk, (3) yields computationally feasible methods for coping with the high complexity in investment analysis, and (4) yields decisions that are optimal rather than satisficing, i.e. performance of the Cybernetic Investment Decision System (CIDS) gradually improves during its operation through learning from past experience. A simplified CIDS has been implemented and tested in actual investment analysis. The experimental results of these tests indicate that through the cybernetic approach quality of investment decisions can be improved.

1. INTRODUCTION

THE RESULTS obtained by conventional methods for stock market analysis to date have not been very successful. The reasons for this are as follows. First, the task of selecting and managing an investment portfolio is a dynamic sequential process. But the time dimension is largely ignored by conventional approaches to portfolio management. The decision problem is clearly a dynamic one and portfolio theory is defective as long as it does not incorporate the time element into its basic structure [7, p. 229].

Secondly, conventional methods rely on techniques developed for decision making under risk. But portfolio management really involves decision making under conditions of *uncertainty* rather than risk because the needed probabilities and statistics are usually initially unknown and cannot be accurately estimated in advance [17, pp. 101–103]. So the design of investment decision systems must be guided by principles for decision making under uncertainty rather than risk.

Thirdly, conventional methods seem to yield many decisions which are satisficing rather than optimal: there is very little analysis of past investment decisions in an attempt to improve on future decisions [14]. In other words, there seems to be very little learning from past experiences with the aim of improving future performance of the investment decision system.

Finally, the stock market (SM) and its environment is a viable system which is exceedingly complex and dynamic—to the extent perhaps of defying a full description. The human mind cannot cope successfully with such complexity, and the situation is thus poorly understood. Clearly, such problems cannot be well handled by statistical tools. But such situations are exactly the type of problems handled by the interdisciplinary science of cybernetics.

Investment analysis is necessarily a complex *judgmental* decision process. It apparently can-

* Assistant Professor of Computer Science at St John's University, Jamaica, New York, and President of Man-Computer Systems, Inc., New York.

not be captured in any simple decision model. So the cybernetic approach strives to automate or program directly the complex judgmental decision processes involved in investment analysis—as they are performed by practitioners in the security industry. But this can be done only with the aid of techniques borrowed from *artificial intelligence*—the advanced form of cybernetics. For example, we may use heuristic programming methods or *perceptron*-type pattern recognition techniques to synthesize the relevant information into investment timing and selection decisions, and machine *learning* algorithms then gradually improve performance of the decision system. The learning mechanism is primarily responsible for the superior performance of the Cybernetic Investment Decision Systems (CIDS).

The practitioners in the security industry make most of their investment selection and timing decisions through determining whether a SM event fits a certain 'pattern'—by adding up evidence derived from measurements or observations of the market and its environment. Specifically, they synthesize investment decisions by weighing evidence derived through an analysis of the underlying economical, political, and monetary fundamentals, through psychological measurements, technical analysis, and so on. A study of this information will reveal patterns and trends in the data which may make prediction possible. But the information patterns needed for successful investment analysis will inadvertently turn out to be so complex that the human mind cannot cope with it. Therefore, all relevant information must be quantified and then artificial intelligence techniques must be used to process it. In other words, the investment decision process must be programmed or automated.

Decision making by weighing evidence can be programmed with the aid of *pattern recognition* techniques. And a powerful mathematical apparatus of pattern recognition (PR) theories can be used in the synthesis of investment decision systems. We shall work with *perceptron*-type PR devices.

The extreme complexity and poor understanding of the situation prevents us from designing optimal investment decision systems initially. But, fortunately, again a powerful mathematical apparatus of *learning system theory* is at our disposal. It enables us to develop effective learning algorithms which will gradually improve the system's performance. So a learning PR scheme is the central information processing element of our Cybernetic Investment Decision Systems (CIDS).

2. CYBERNETIC INVESTMENT DECISION SYSTEMS

The investment decision problem is far too complex and cannot therefore be handled by any simple problem-solving procedure. So we use the hierarchical approach to problem solving. The spirit of this approach is to divide and conquer: break up a complex problem into simpler subproblems that are within the reach of our capabilities, develop solutions (or problem-solving systems) for these subproblems, integrate them into a viable system, and thus come out with a solution to the original problem.

Accordingly, our CIDS is hierarchically structured. We factor the global investment decision problem into two major subproblems—investment *timing* and *selection*. The timing problem is further factored into two local subproblems: (A) general market timing, and (B) individual stock transaction timing. Similarly, the selection problem is also factored into two local subproblems: (C) security analysis, and (D) efficient portfolio diversification. This hierarchical structure of the investment decision process is diagrammed on Fig. 1.

Subproblems (A), (B), and (C) involve decision making by weighing evidence. For example, general market timing decisions (i.e. stock market forecasts) are made by weighing evidence obtained through and analysis of the market's (monetary, political, and economical) fundamentals, technical factors, psychological measurements, observations of the news back-

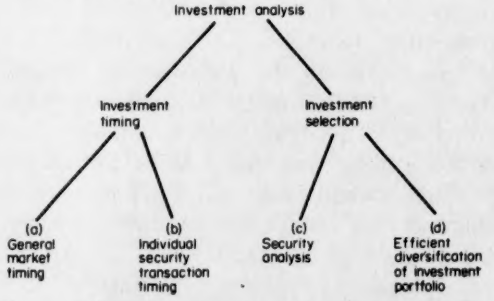


FIG. 1. *Hierarchical structure of the investment decision process.*

ground, and so on. Similarly, investment selection (security analysis) decisions are synthesized from observations of earnings, dividends, earnings trends, price earnings ratios; from technical analysis, through measurements of the underlying psychological factors, etc. Thus subproblems (A), (B) and (C) can be programmed with the aid of pattern recognition techniques. Subproblem (D) can be handled by mathematical (quadratic) programming methods and therefore will not be considered any further.

To each local subproblem (A), (B), or (C) corresponds a separate decision subsystem of the CIDS. And each subsystem is again hierarchically organized: It is designed in form of a functional three-layer hierarchy so that higher layers determine or adjust some parameters on the lower levels.[1] That is, the system can be *functionally* decomposed into three layers: (1) a layer of programmed decision processes, (2) a learning layer, and (3) a layer of non-programmed decision processes (global planning layer). This functional hierarchy emerges naturally in reference to three essential aspects of the investment decision process: (1) the search for a preferable or acceptable course of action under known, prespecified conditions; (2) the reduction of uncertainties and improvement of performance; and (3) the selection of strategies to be used in investment management. The flow of information and control between these three functional parts of the CIDS is displayed on Fig. 2.

Each level of the decision system performs a different function. The first layer represents all those aspects of the investment decision process that can be automated, i.e. it denotes formal (mechanical) 'systems' for security analysis and market timing. The function of the second layer is to increase the knowledge about the system and its SM environment through a learning process with the aim of improving performance of the decision system after it has been implemented and put into operation. The functions of this layer can either be automated or realized by the human operator. The third layer represents all those activities related to investment analysis that cannot be readily formalized and mechanized on the lower layers. This includes (1) design and subsequently modify (if necessary) lower layers of the decision system, and (2) perform those tasks of the investment decision-making process that have not been automated on the lower layers. Of course, the third layer functions are performed by man.

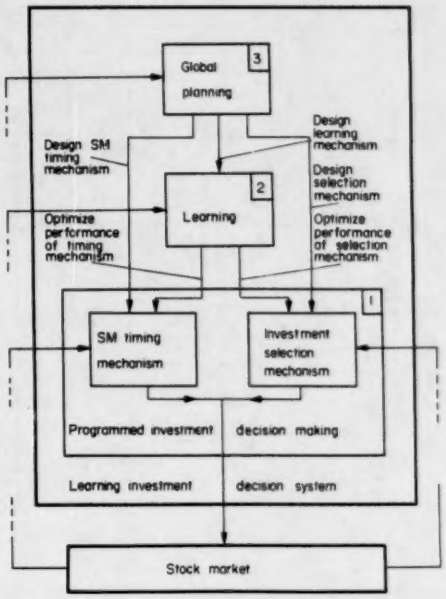


FIG. 2. *Functional multilayer hierarchy of an investment decision system.*

2.1 *General approach to programming investment decisions*

We'll develop the CIDS within the conceptual framework of a general model for automation of decision making and problem solving [4, 5]. This model factors the decision process into four phases: (1) the *identification* activity which consists of searching the environment and numerically encoding information for investment analysis. (2) Then in the *design* phase alternative solutions to the investment decision problem are developed. (3) The best among the alternative solutions is then selected and implemented in the *choice* phase. And (4) through a subsequent evaluation of the implemented solution *learning* feedback is generated. Hence, the anatomy of this decision process can be diagrammed by the four stage information *feedback system* on Fig. 3.

In the following subsections we now outline an application of learning pattern recognition

[1] For a more thorough discussion of such problems, see the book of Mesarovic, *et al.* [10].

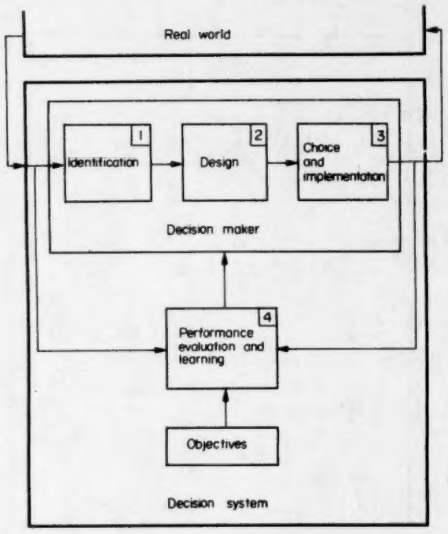


FIG. 3. *The anatomy of the learning decision process.*

techniques to programming the individual phases of the investment decision process.

The identification activity. The decision situation is numerically encoded by the *state vector* $\mathbf{x} = (x_1, \ldots, x_r)$. We define the state of a system and its environment at any time as the information needed to determine the behavior of the system from that time on. Hence the state variables x_i, $i = 1, \ldots, r$, represent observables or measurements characterizing the various features (properties or attributes) of the investment decision situation that have some predictive significance.

Since the amount of computation needed to process an information pattern grows approximately exponentially with its size, through a mathematical *Feature Space Transformation* (FST) the state vector $\mathbf{x}$ is transformed into a *feature* vector $\boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_s(\mathbf{x}))$. This FST greatly reduces the size of $\mathbf{x}$ but without reducing too much its information content. That is, the FST performs aggregation, filtering, and condensation of investment information, so as to emphasize those aspects of the situation that are deemed important.[2] In short, $\boldsymbol{\phi}(\mathbf{x})$ is a numerical encoding of the information for investment analysis and is the input for the second phase of the decision process.

Design and choice of the optimal investment policy. Decision making is the process of converting information into action, i.e. it is a mapping from the state (or feature) space onto a policy space. When programming *decision making by weighing evidence*, we can realize this mapping by *perceptron*-type pattern recognition techniques.

Let the investment policy space be $\mathbf{A} = \{a_i\}$, where a_i, $i = 1, 2, 3, \ldots$, are the alternative investment policies, e.g. buy, sell, hold, sell short, etc. and let's define linear *discriminant* functions

$$g_i(\mathbf{x}) = \mathbf{W}_i \cdot \boldsymbol{\phi}(\mathbf{x}) = \sum_{k=1}^{s} w_{ik} \cdot \phi_k(\mathbf{x}) \qquad (1)$$

where $\mathbf{W}_i = (w_{i1}, \ldots, w_{is})$ is the weight vector.[3] Then the choice mechanism (i.e. the mapping from investment information into a course of action) can be implemented by a set of discriminant functions. Thus we associate a unique discriminant function with every investment policy a_i (except if the policy space has only two courses of action, then one discriminant suffices). This is expressed in mathematical terms by

$$a_i \leftrightarrow g_i(\mathbf{x}).$$

Then for any investment decision situation the system selects that policy a_i for which the corresponding $g_i(\mathbf{x})$ is highest, i.e.

$$g_i(\mathbf{x}) > \max_{i \neq j} g_j(\mathbf{x}). \qquad (2)$$

In this way a unique investment policy is associated with every pattern of information $\mathbf{x}$.

If the policy space contains only two courses of action, we need only one discriminant $g(\mathbf{x})$. Then a_1 is selected if $g(\mathbf{x}) > 0$, and a_2 is chosen otherwise.

The learning mechanism. The purpose of the learning process is to gradually adjust certain system parameters so as to optimize the decision system's performance. The learning process is directed by information feedback derived through an evaluation of past investment decisions. The performance criterion is always some measure of the expected return and risk. It can be chosen in several ways. For our present purpose, we use the probability of error as the performance index. By 'error' we mean that a wrong investment (timing or selection) decision has been made, e.g. the expected out-

[2] See other works of this author [4, 5] for the technical and mathematical details of the identification phase.

[3] A good description of discriminant techniques in pattern recognition can be found in Nilsson [12].

come was not attained. The decision system is then optimized by minimizing the probability of error, i.e. the average number of errors. Of course, minimizing the expected number of wrong trading decisions is equivalent to maximizing the expected return on investments.

The performance of the decision system is measured by a performance evaluation function $V(\mathbf{W},\mathbf{x})$. Usually $V(\mathbf{W},\mathbf{x})$ represents some measure of error. The expected value of $V(\mathbf{W},\mathbf{x})$ is the *performance index* $I(\mathbf{W})$ of the decision system; that is

$$I(\mathbf{W}) = E\{V(\mathbf{W},\mathbf{x})\}.$$

And the objective of the learning process is to find that setting of system parameters $\mathbf{W}^*$ for which the functional $I(\mathbf{W})$ takes its extreme value, i.e. for which

$$\text{grad } I(\mathbf{W}) = E\{\text{grad } V(\mathbf{W},\mathbf{x})\} = 0.$$

where the *gradient* operation is taken with respect to the weight vector $\mathbf{W}$.

Learning system theory [15] provides us with the following recursive algorithm for iteratively determining the optimal weight vector $\mathbf{W}^*$:

$$\mathbf{W}_{n+1} = \mathbf{W}_n - (a_0/n)\text{grad } V(\mathbf{W}_n,\mathbf{x}_n), \quad (3)$$

where a_0 is a constant and the subscript n represents the *time* when $\mathbf{x}$ and the corresponding $V(\mathbf{W},\mathbf{x})$ was observed or measured, i.e. $\mathbf{W}_n$ and $\mathbf{x}_n$ are computed at n discrete time intervals.

Let us assume for simplicity that our policy space contains only two courses of action, i.e. $\mathbf{A} = \{a_1, a_2\}$. Also let us define

$$V(\mathbf{W},\mathbf{x}) = [y - \text{sign } g(\mathbf{W},\mathbf{x})]\cdot g(\mathbf{W},\mathbf{x}) \quad (4)$$

where $y = 1$ if a_1 was the correct decision and $y = -1$ otherwise, and sign $z = 1$ if $z > 0$ and sign $z = -1$ otherwise. (The value of y represents the feedback information determined by the performance evaluation of actually made investment decisions after they were carried out. We also observe that $V(\mathbf{W},\mathbf{x})$ has a nonzero value only when the decision system made an error. Thus $I(\mathbf{W})$ is proportional to the error-rate and its minimization by algorithm (3) will reduce the decision system's probability of error [15]. Then substituting (4) into (3) we obtain the learning algorithm

$$\mathbf{W}_{n+1} = \mathbf{W}_n + (a_0/n) \times [y - \text{sign } g(\mathbf{W}_n,\mathbf{x}_n)]\cdot\phi(\mathbf{x}_n). \quad (5)$$

Figure 4 shows the general structure of a two-category learning PR system. We shall see later that scheme (5) provides us with a computational procedure or program for gradually determining that configuration of system parameters $\mathbf{W}^*$ for which the investment decision system's performance is best (i.e. $I(\mathbf{W})$ is minimum), that is, the probability of making the wrong investment decision is smallest.

Next we will apply the above concepts to programming investment timing and selection.

2.2. *General market timing*

In this article we shall consider only an application to forecasting the general market, i.e. timing subproblem (A). The application to in-

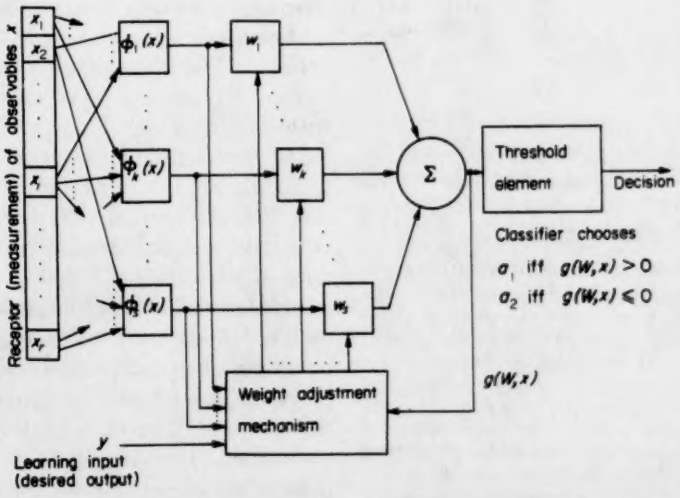


FIG. 4. *Structure of a perceptron-like learning pattern recognition system.*

dividual security transaction timing (subproblem B) will be considered in our other publications.

With respect to general market timing, we may define the general market *state* in terms of stock price *trends* and trend reversals. Many different states can be defined, but we find it convenient to work with four states of the market: *uptrend*, *top*, *downtrend* and *bottom*.

Now, many market practitioners are trend followers: they prefer to *buy* at major market bottoms, *hold* during the uptrend, *sell* (or sell short) near a market top, and *do nothing* (or hold short positions) during a downtrend. Thus trading in the stock market is a *recurrent* decision process whose policy space contains four well-defined courses of action. And we can program this decision process with the aid of PR techniques.

The success of the programmed decision system depends on its ability to predict future direction and changes of trends in stock prices. It is apparently impossible to predict future price changes through an analysis of past price changes. But if the information pattern for stock market analysis becomes *very complex*—and includes also fundamental state variables, psychological measurements, news background, etc. in addition to technical factors—then prediction of future price changes appears possible. The information pattern **x**, which serves as a basis for investment decisions, must necessarily be so complex that the human mind is generally unable to ascertain the correlation between present values of **x** and future changes in stock prices. But this relationship can be gradually determined through a machine learning algorithm.

We shall discuss in detail the application of learning PR techniques to stock market prediction in Section 3. But first let us consider an application to security analysis.

2.3. *Programming investment selection*

The design of every programmed investment selection system must be tailored to requirements and conditions existing within individual user organizations (e.g. investment research departments). Thus the first step in the design process is to specify the user's investment objectives in operational terms, i.e. a *selection criterion* must be established. Within the decision system these investment objectives are represented by a performance index.

There are many possible selection (performance) criteria. The relative strength criterion is often used. So let us use it as an example throughout this work. We define a security to be relatively strong over a given time horizon if it appreciates more than a general market index during that time while entailing no more than average (market) risk, and it is relatively weak otherwise. Clearly, relatively strong issues are characterized by a different information pattern than relatively weak stocks. In other words, this selection criterion divides the set of stocks into two classes: those that are in a relatively strong state and those in a relatively weak state (during a prespecified time horizon). So a two-category pattern classifier can be used to program investment selection according to this criterion.[4]

The second important design step is developing the mechanism for identification of information for decision making in form of the *state vector* **x**. The state variables in **x** represent a numerical encoding of various properties and observables of individual securities. They include fundamental characteristics, e.g. earnings trend, earnings stability, future earnings potential, dividend yield, dividend trend, price earnings ratio, etc.; risk characteristics, i.e. the variance of past price changes or the Beta coefficient; technical factors like current price trends on the charts, volume measurements, price overhang, etc. psychological factors, e.g. divergence measurements; and various other factors like news background, capitalization, quality of institutional sponsorship of the security, company's reputation for technical brain-power and research, quality of management, Standard

[4] We have not considered the element of risk. This was done on purpose for several reasons. First, the commonly accepted measures of risk, i.e. the Beta coefficient or variance of past price changes, are nonstationary and thus are not useful measures of risk. Besides, these quantities are not always available for many securities, like, for example, new issues. Secondly, and more importantly, investment decisions are made under conditions of uncertainty rather than risk [17], and the proper way to deal with uncertainty is to try to reduce it by increasing our knowledge and understanding of the situation, e.g. through a learning process. But we can resolve most problems of this type by limiting our selections to a prespecified range or level of risk as measured, for example, by the variance of past price changes, if it is available. Thus in our tests we assume we work with a portfolio of average (market) risk, e.g. Beta = 1.

& Poor's rating of the security, and so forth.

The state vector $\mathbf{x}$ is then mathematically transformed into a pattern of relevant features $\phi(\mathbf{x})$ which represent the corresponding fundamental, technical, psychological, and other features of individual securities. This FST is designed with the aim of greatly reducing the size of $\mathbf{x}$, but without also reducing too much of its information content.

Next we shall program the decision rule. Let the index n count the number of securities considered for investment. Then the selection rule can be programmed by discriminant functions:

select the n-th security if $g(\mathbf{x}_n) > 0$, and
discard it otherwise.

This rule will associate a selection policy, e.g. buy (hold) or sell (discard), with any pattern of information $\mathbf{x}$ describing the security. (With this scheme we can also program the individual security's transaction timing (Subproblem B) using methods similar to those described in the next section.)

Finally, the selection rule is optimized through a learning process with the aim of minimizing the probability of error. The learning process functions this way. At the beginning we set the weights $\mathbf{W}$ according to our best initial knowledge about what features are known to be important; or we may pick the weights arbitrarily, e.g. set them all equal to one. Afterwards the weights are changed under the direction of error-correcting feedback derived through performance evaluation of actually made past stock selections. For example, suppose that the n-th security has been selected (bought or held), that is $g(\mathbf{x}_n) > 0$. If this stock subsequently outperforms the market during the prespecified time horizon, the system made a correct decision. (In this case the weight vector remains unchanged but the counter n is increased by one.) But if the security appreciated less than the market index, a wrong decision was made, and the weight vector $\mathbf{W}_n$ is subsequently changed according to algorithm (5). This 'correction' reduces the chance of making a similar poor selection in the future. The system is also 'corrected' when a reverse error was made, i.e. when a relatively strong stock was not selected. The new value of the weight vector $\mathbf{W}_{n+1}$ is then used in subsequent computations until the next error is made.

After some time, this step-by-step learning from past experiences may result in real performance improvement. At the start, the system's performance will most likely be only mediocre. But after about twenty learning steps ($n = 20$) the system may become close to optimal and may make better decisions than a human analyst could do intuitively. (See our experimental results in Section 4.)

3. AN ACTUAL IMPLEMENTATION

We have implemented a cybernetic general market timing mechanism. It has been experimentally tested in *actual* investment analysis since 1970.

Our scheme may be regarded as an attempt to program the well known General Indicator Approach to market timing. This method synthesizes investment decisions by weighing evidence obtained through an analysis of a broad spectrum of *indicators*, i.e. measurements or observations, describing the state of the market.

We observe that the CIDS is a man-machine system which is tailored to the needs of its users. So its design will generally reflect the user's resources like the type of information to which he has access, his attitudes toward risk and expected returns on investments, available computational resources, etc. For these reasons *there is an infinity of possible realizations of CIDS*. The choice of our realization has been guided primarily by the requirement of computational simplicity. In fact, the present investment decision system is so simple that all required computations can be efficiently performed by hand.

The decision system again operates in four phases. The first phase, the *identification* activity, consists of searching the environment and numerically encoding all information characterizing the state of the general market. The relevant information is obtained by monitoring the various fundamental, technical, economical, monetary, and psychological indicators; reading daily newspapers; watching the investment news background, and so forth. All these observations and measurements are represented by the state vector $\mathbf{x} = (x_1, \ldots, x_{51})$. So the state variables x_i, $i = 1, 2, \ldots, 51$, represent a numerical encoding of individual SM indicators, observations of the news background,

psychological measurements, etc. In short, x contains all the information about the market's past that is relevant for the prediction of its future. The encoding of state variables is done objectively and subjectively.[5]

Through an FST the 51 state variables in x are transformed into feature vectors representing the fundamental, technical, psychological, etc. features of the general market. Now, different investment policies are generally used for different time horizons. And different information patterns are needed to predict the market's behavior during different time horizons. Thus, since we'll work with three time horizons—short, intermediate, and long-term—we'll need three feature vectors $\phi_i(\mathbf{x})$, $i = 1, 2, 3$, respectively. The elements of $\phi_i(\mathbf{x})$ represent a numerical encoding of the following general market features:[5]

- $\phi_{i1}(\mathbf{x})$—aggregation of political, economical, and monetary *fundamentals*.
- $\phi_{i2}(\mathbf{x})$—*technical factors*, e.g. advance-decline line, volume and price trends, etc.
- $\phi_{i3}(\mathbf{x})$—aggregation of *psychological measurements* ascertaining the present attitudes of the investing public.
- $\phi_{i4}(\mathbf{x})$—a quantitative measure of the *news background* underlying stock price changes.
- $\phi_{i5}(\mathbf{x})$—measurements derived from the *Elliot Wave Theory*.
- $\phi_{i6}(\mathbf{x})$—market statistics like average duration and extent of SM price tends.
- $\phi_{i7}(\mathbf{x})$—a measure expressing the present market *efficiency*.

The short-term feature vector ($i = 1$) contains five of the above features, the intermediate-term vector ($i = 1$) contains five of the above features, the intermediate-term vector ($i = 2$) contains all seven features, and the long-term vector ($i = 3$) contains six features. All information needed for the SM identification activity is obtained from the Wall Street Journal and BARRON'S.

[5] The inclusion of techniques for numerically encoding the state variables x_i, $i = 1, \ldots, 51$, and mathematical formulas for computing the features; ϕ_{ij}, $i = 1, 2, 3$, and $j = 1, \ldots, 7$, would greatly increase the size of this presentation. These schemes are described in our other works [4, 5]. A method for numerically encoding the investment news background can also be found in a paper by Niederhoffer [11].

The design and choice mechanism, i.e. the mapping from the decision situation onto the policy space, is again realized by discriminant functions. Assume that our policy space contains four courses of action, e.g. buy, hold, sell (sell short), and do nothing (hold short positions). Since we need a policy space for each time horizon, we shall work with three policy spaces: $\mathbf{A}_i = \{a_{ij}\}$, $i = 1, 2, 3$, and $j = 1, \ldots, 4$. For example, a_{12} then represents short-term hold, a_{31} means long-term buy, etc.

We program the choice mechanism by associating a unique discriminant with every trading policy:

$$a_{ij} \leftrightarrow g_{ij}(\mathbf{x})$$

where

$$g_{ij} = \mathbf{W}_{ij} \cdot \phi_i(\mathbf{x}).$$

Then for any investment decision situation the system selects that set of trading policies a_{ij}, $i = 1, 2, 3$, for which the corresponding

$$g_{ij}(\mathbf{x}) > \max_{j \neq k} g_{ik}(\mathbf{x}), \quad j, k = 1, \ldots, 4.$$

Performance of this choice mechanism is then gradually optimized through a learning procedure. This learning decision making process is summarized by the flow-chart on Fig. 5.

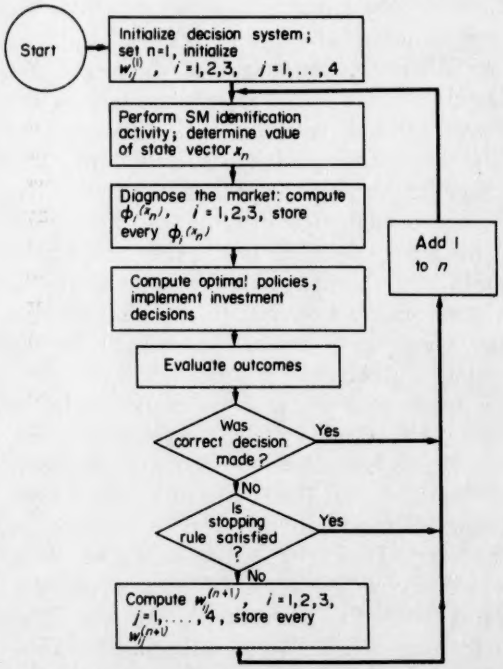

Start
Initialize decision system; set n=1, initialize $w_{ij}^{(1)}$, $i = 1,2,3$, $j = 1, \ldots, 4$
Perform SM identification activity; determine value of state vector x_n
Diagnose the market: compute $\phi_i(x_n)$, $i = 1,2,3$, store every $\phi_i(x_n)$
Compute optimal policies, implement investment decisions
Evaluate outcomes
Was correct decision made?
Yes
No
Is stopping rule satisfied?
Yes
No
Compute $w_{ij}^{(n+1)}$, $i = 1,2,3$, $j = 1, \ldots, 4$, store every $w_{ij}^{(n+1)}$
Add 1 to n

FIG. 5. *Flow diagram of the learning algorithm for trading in the stock market.*

To simplify the computations, we reduce the size of our three policy spaces to two courses of action: (1) *buy* (cover shorts) or *hold* long positions if bought earlier, and (2) *sell* (sell short) or *do nothing* (hold short positions) if sold earlier. In this case we need only three discriminant functions—one for each time horizon. Then, for example, a short-term trading strategy can be mechanized by the following program:

(1) *buy* (cover shorts as soon as $g_1(\mathbf{x}) > 0$, after a decline, then
(2) *hold* as long as $g_1(\mathbf{x}) > 0$, then
(3) *sell* (sell short) as soon as $g_1(\mathbf{x}) \leq 0$, then
(4) *do nothing* (hold short positions) as long as $g_1(\mathbf{x}) < 0$, and when $g_1(\mathbf{x}) \approx 0$, go to (1).

And through the application of *heuristic* methods we can also program more complex trading strategies involving more than one time horizon.

The error-rate, i.e. the average number of wrong investment decisions, is then minimized through a learning process. The learning mechanism is initialized as follows. The three time counters $\mathbf{n}$ are set to one. (We note that for the short-term decision subsystem a unit of time is 1 week, for the intermediate term the unit time interval is 4 weeks, and for the long term it is 16 weeks.) The weighing coefficients a_0 are set to one, and all weight vectors $\mathbf{W}_i$, $i = 1, 2, 3$ are initially set to one.

The learning decision process proceeds gradually in an iterative manner. The process starts with the identification activity which is carried out continuously. Thus every day the Wall Street Journal is read and other news media are scanned for information that may be relevant to the stock market. Any relevant information is recorded and later encoded in the state variables x_i, $i = 1, \ldots, 51$. At regular time intervals, (e.g. every week for short-term analysis, every 4 weeks for intermediate-term analysis, etc.) the state vector is transformed into feature vectors $\boldsymbol{\phi}_i(\mathbf{x})$, $i = 1, 2, 3$, and the corresponding set of trading policies is computed. These policies are then implemented and the results are observed and evaluated. If the evaluation of results indicates that the optimal policy has been followed, no corrective actions are taken. However, if it later becomes clear that a wrong investment policy has been carried out, the decision system is corrected according algorithm (5). For instance, suppose that trading policy a_{21} (intermediate-term buy or hold) is carried out at time $\mathbf{n}$, that is, $g_2(\mathbf{x}_n) > 0$; but during the next several weeks the market suffered an intermediate sized decline, which resulted in sharp decline in value of the investment portfolio. Consequently, the value of $\mathbf{W}_2$ is 'corrected' by computing

$$\mathbf{W}_2^{(n+1)} = \mathbf{W}_2^{(n)} - (1/n)\boldsymbol{\phi}_2(\mathbf{x}_n), \tag{6}$$

and the value of $\mathbf{W}_2$ is used in subsequent computations of the trading policy. If the market would not have declined, the weight vector would remain unchanged (i.e. $\mathbf{W}_2^{(n+1)} = \mathbf{W}_2^{(n)}$), but the time counter n would be increased by 1 four weeks later. The same correction would take place if the a_{22} policy would be in error, but the sign in equation (6) would be reversed. The weight vector remains unchanged until the next error is made. This error–correction learning process continues until the error rate becomes stationary.

We note that the same learning algorithm is employed for all three time horizons, but the required learning times differ approximately by a factor of four between time horizons. Thus, if it takes on the average 20 weeks to optimize the short-term decision system, it may take 80 weeks to optimize the intermediate-term system, and it might take approx 6 years to optimize the long-term decision system. However, methods for acceleration of the learning process do exist [5, Ch. 3].

4. EXPERIMENTAL RESULTS

We have implemented a CIDS designed according the principles outlined in this paper and tested them in *actual* investment analysis since 1970. The experimental results of these tests are encouraging: they seem to indicate that with the cybernetic approach we can (1) program some important judgmental aspects of investment analysis; (2) improve the quality of investment decision making, i.e. amplify the intellect of the human analyst; and (3) attain above average investment performance.

4.1 *General market forecasting*

We have tested two mechanical trading strategies designed according to principles outlined in Section 3. In the first of them (S-1) selling short is not permitted, i.e. during declining markets a 'do nothing' policy is followed. In

the second strategy (S-2) selling short is permitted. An intermediate-term time horizon has been used. That is, all decisions are made at intermediate market turning points, i.e. when the sign of $g_2(\mathbf{x})$ changed.

The two major results of our experimental testing are summarized below:

(1) Performance of the decision system was gradually improved by the learning algorithm. The improvement was considerable. For instance, during a 20-week period the error-rate of the short-term timing mechanism has dropped from almost 50% to approx 20–30%. Similar improvement of performance was observed on the intermediate- and long-term decision systems. The same learning algorithm is employed for all three time horizons, but the required learning times differ approximately by a factor of four between time horizons. The short-term learning curve is shown on Fig. 6. During a 49-week time period the short-term weight vector has changed from the initial value $\mathbf{W}_1^{(1)} = (1, 1, 1, 1, 1)$ to $\mathbf{W}_1^{(49)} = (1.307, 1.145, 1.045, 0.850, 0.202)$.

(2) After some learning period, reversals of stock market trends were usually recognized sufficiently early for profitable action. The relationship between the values of discriminants and the corresponding SM trends is displayed on Fig. 7. It is seen that near the start of an uptrend the value of the corresponding discriminant becomes positive, it remains usually positive during the entire duration of the uptrend, and then it becomes negative near the beginning of the next downtrend. (But, of course, there are some errors from time to time.) Thus the trading policies generated by the decision system appear to result in above average returns while taking no more than average risks. For example, from 1970 to 1973, strategies S-1 and S-2 were applied to the NYSE Composite Index. These strategies generated net (after commissions) returns of 32.5% and 38%, respectively, while the general market

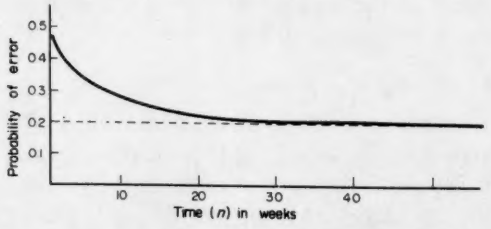


FIG. 6. *Learning curve of the short term decision subsystem.*

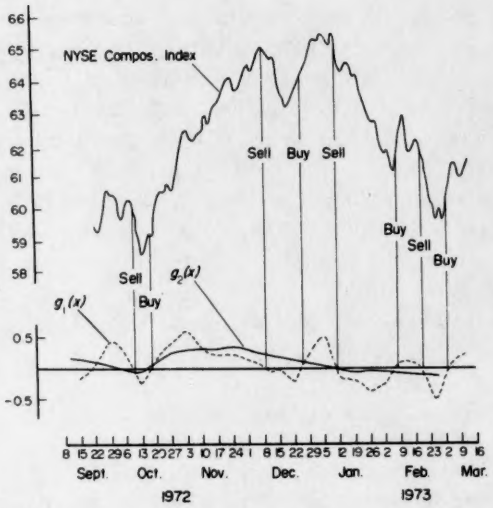


FIG. 7. *Values of short- and intermediate-term discriminants* ($g_1(x)$ *and* $g_2(x)$, *respectively*) *from September* 1972 *to March* 1973 *with subsequent changes of stock prices.*

return during the same period was 29% (dividends were discarded). A commission of 2.5% per round trip (in and out) trade has been assumed. The risk of strategy S-1 is below average because approximately one-third of the time investible funds are held in riskless cash, while the risk of strategy S-2 is average.

4.2 *Automating investment selection*

We have implemented a simple programmed scheme for security analysis designed according to the principles outlined in Section 2.3. The relative strength performance criterion has been used.

The results attainable by such programmed selection systems appear to be indeed superior to intuitively made selections. For example, we have tested this scheme by participating in The 1972–1973 Value Line Contest of Stock Market Judgment. The 25-stock portfolio selected by the CIDS was placed 342nd out of a total of 89,744 entries. Within a 6 months period it outperformed the averages by almost 14% while taking no more than average risks, and it won a cash prize.

It must be pointed out, however, that these results should be considered as preliminary because our experiences with the cybernetic approach are rather limited to date. More research into this approach, and more experimental testing of our CIDS will be performed in the future.

5. SUMMARY AND CONCLUSIONS

The cybernetic approach derives its strength from three sources. First, it is a *learning approach*. This means that CIDS can be designed and implemented without a thorough initial understanding of the investment decision situation. The system then compensates for our initial lack of knowledge by gradually improving its performance during its operation—using *current information*—as new information becomes gradually available.

Secondly, CIDS are viable systems because they are complex. They are (at least theoretically) able to synthesize investment decisions from *all* relevant information. In other words, this approach conforms with the Cybernetic Law of Requisite Variety which is so essential for SM success.

The third source of strength derives from *synergism* in complex man–machine systems. That is, the cybernetic approach to SM analysis is an integrated system whose components are many conventional investment methods: the fundamental approach, various technical methods, psychologically oriented techniques, as well as some aspects of conventional capital market theory, are incorporated into the CIDS (as features) and thus are dynamically interacting subsystems of the cybernetic approach. The CIDS is then so designed that the simultaneous interaction of the individual methods (under direction of a learning process) yields better results than would correspond to the sum of the individual SM techniques taken independently of each other. In other words, the CIDS is more than the sum of its parts.

In conclusion, we re-emphasize that investment analysis is an extremely complex judgmental decision process which apparently cannot be successfully studied within the framework of any single discipline like, for instance, probability theory and statistics. An interdisciplinary approach is needed. And we believe that the interdisciplinary science of cybernetics is well qualified to provide a conceptual framework for investment analysis. With techniques borrowed from cybernetic disciplines we can probably obtain better results than with purely statistical tools. In fact, the cybernetic approach may become the vanguard of future directions in this field. Of course, this is a tentative hypothesis which must be thoroughly tested. Therefore, this research will be pursued further. But it involves much more work than one man can handle alone. So, hopefully, this work will stimulate others to participate.

REFERENCES

1. Beer S (1966) *Decision Control*. John Wiley, London.
2. Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *J. Finance* 383–417.
3. Felsen J *Low-Cost, Personal-Computer-Based Investment Decision Systems*, (to be published).
4. Felsen J (1975) *Cybernetic Approach to Stock Market Analysis versus Efficient Market Theory*. Exposition Press, Hicksville, NY.
5. Felsen J (1976) *Decision Making under Uncertainty: An Artificial Intelligence Approach*. CDS, New York.
6. Friend I, Blume M & Crocket J (1970) *Mutual Funds and Other Institutional Investors: A New Perspective*. McGraw-Hill, New York.
7. Granger CWJ & Morgenstern O (1970) *Predictability of Stock Market Prices*. DC Heath, Lexington, MA.
8. Hayes JR (1962) Human data processing limits in decision making. Report No. EDS-TDR-62-48, Air Force System Command, Bedford, MA.
9. Jensen ME (1972) The foundations of the current state of capital market theory in *Studies in the Theory of Capital Markets* (Ed. Jensen MC). Praeger, New York.
10. Mesarovic MD, Macko D & Takahara Y (1970) *Theory of Hierarchical, Multilevel Systems*. Academic Press, New York.
11. Niederhoffer V (1971) The analysis of world events and stock prices. *J. Bus.* (April) 193–217.
12. Nilsson NJ (1965) *Learning Machines*. McGraw-Hill, New York.
13. Simon HA (1960) *New Science of Management Decision*. School of Commerce, New York University, NY.
14. Smith KV (1971) *Portfolio Management*. Holt, Rinehart & Winston, New York.
15. Tsypkin YS (1971) *Adaptation and Learning in Automatic Systems*. Academic Press, New York.
16. Vasicek OA & McQuown JA (1972) The efficient market model. *Fin. Analy. J.* (Sept–Oct).
17. Williamson JP (1971) *Investments: New Analytical Techniques*. Praeger, New York.
18. Wyckoff P (1963) *The Psychology of Stock Market Timing*. Prentice-Hall, Englewood Cliffs, NJ.

ADDRESS FOR CORRESPONDENCE: *Jerry Felsen Esq, Department of Mathematics and Computer Science, St John's University, Grand Central and Utopia Parkways, Jamaica, New York* 11439, *USA.*

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 249–256
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0249$02.00/0

Short-Term Foreign Exchange Risk Management: Zero Net Exposure Models

MARTHA S HOLLIS

University of California, Irvine

(*Received July 1977; in revised form October 1977*)

By balancing the relevant end of period assets and liabilities (i.e. a zero net exposure approach to foreign exchange risk management) the multinational corporate treasurer can, conservatively, protect against parity fluctuations without requiring exchange rate forecasts. This paper presents two goal programming formulations for short-term money management and evaluates the resulting strategies (via simulation) in terms of their impact on corporate profits. A hypothetical multinational corporation with headquarters in the US and subsidiaries in Canada, West Germany, and the United Kingdom serves as the experimental unit. The results suggest that the zero net exposure approach is no better than a naive 'dart-throwing' approach.

THE RESULTS of short-term money management decisions can be influenced by foreign exchange rate fluctuations. Exchange rate movements can lead to currency gains and losses that are reflected in the corporation's net income statement. To offset this foreign exchange rate risk some analysts have suggested using a zero net exposure approach to short-term money management (that is, the balancing of end-of-period short-term monetary assets and liabilities).[1] Prior work on net exposure models has been predominately qualitative in nature and can be characterized by the omission of specific details such as the length of time horizons, how variables are related, and the mathematical form for finding solutions.[2] Qualitative methods, because of their lack of specificity in dealing with the multitude of factors which are of vital importance in the decision-making process, cannot demonstrate the quantitative impact on profits. This research overcomes some of the deficiencies of prior studies in the international money management process by stating precisely two of the most frequently advocated zero net exposure approaches in the form of mathematical models.

Two archtypical models of multinational corporations' (MNCs) organizational control are explored: a centralized single period zero net exposure model (CSP) and a decentralized multiperiod zero net exposure model (DMP). The third model, developed for comparative purposes, represents a naive random selection process (NRS). Each model is used to obtain monthly strategies for a hypothetical US MNC operating during the time period September 1974 through August 1975. The twelve monthly strategies that are obtained from each of the three models are submitted to a simulator that determines what would have happened with respect to before-tax consolidated profits if the transactions had been implemented during this period of volatile exchange rate fluctuations. Profits resulting from simulated implementation are compared using nonparametric inferential statistics. This research should provide

[1] Zero net exposure is defined herein as the balancing of end-of-period short-term monetary assets and liabilities to avoid rate risks eschewing all types of forward contractual arrangements and other hedging instruments. For other definitions of net exposure see Teck [14].

[2] See, for example, the expositions of Eiteman and Stonehill [2], Goeltz [5], Korth [9], and Teck [14].

MNCs with some preliminary guidance regarding their short-term money management decision-making. This paper suggests that there are no significant differences among the profits resulting from the strategies suggested by the two zero net exposure models and a naive random approach.

The assumptions of net exposure models are detailed in Section I. The mathematical formulations for the centralized single period model, the decentralized multiperiod model, and the naive random strategy are presented in Sections II, III, and IV, respectively. Profit results of the simulated applications are presented and tested in Section V. The final section contains conclusions and areas for future research.

I. NET EXPOSURE ASSUMPTIONS

In the CSP and DMP models, to preclude the necessity for forecasting directions of exchange rate movements, the firm is assumed to strive for an ending exposure balance as close to zero as possible in each country of operation.[3] The firm's exposure position is assumed to preclude any willingness to experience exchange rate fluctuations.

This study is limited to the essential decision variables of short-term financial management in multinational corporations. These variables, derived from prior work in the theory of financial management, are: investments with duration of less than one calendar year and borrowings with repayment terms less than one calendar year.[4] Short-term investments are included by considering only the most active classification in each country (those similar in risk and liquidity to the US Treasury Discount Bill). Borrowing opportunities will be restricted to commercial banking operations in each country.

Short-term financial decisions are made using a multitude of period lengths. The strategies studies here allow the firm to select alternatives and implement decisions on a monthly basis (based on a total horizon of either one or three periods in the future) over the course of one year. The period of one calendar year is chosen to demonstrate the total results of short-term transactions which would be summarized on MNC's consolidated annual financial statements. Presently, many firms operating in the international markets consider such decisions on a quarterly basis to coincide with forecasts provided by large commercial banks' multinational money management centers. Given the anticipated increased parity fluctuations, a monthly basis appears to be more advantageous as it affords an opportunity to update information on a more timely basis. It is infeasible to consider a shorter time period (weekly or daily) due to the lack of available data.

The following assumptions are made in order to limit the study of the short-term financial problems to the use of net exposure models:

1. The cash budgets for the subsidiary in each country, for each monthly period, are exogenously determined;

2. The cost (return) and availability of financing and investment opportunities are known, as are the firm's upper limits on such activities;

3. Commissions, transaction costs, and other compensatory charges are not considered;

4. The firm's transactions in various markets are not of significant magnitude to influence the various rates and costs of financial instruments, that is, the firm's operations are independent of the markets' rate determination system.

Assumption 2 represents the only departure from the condition that data used to obtain a strategy at any point in time would be available to a decision maker. A decision maker would not know the exact value of future borrowing and investment rates. When such specifications become pertinent to decision making, forecasts will be required.

[3] See Giddy and Dufey [4] for recent work on exchange rate forecasting techniques and their results.

[4] Other short-term variables include the management of inventories, accounts payable, and cash management efficiency movements (such as lock boxes and wire transfers). These are not considered in order to concentrate on purely monetary decisions. Hedging devices are excluded to focus attention on pure net exposure approaches.

Vol. 6
1978

II. CENTRALIZED SINGLE PERIOD (CSP) MODEL

The first model represents a centralized single period, CSP, approach to the money management problem.[5] Located in the domestic country (United States) is a cash pooling center serving as a medium for transferring funds among the countries in which the firm has operations. Decision-making is centralized at the domestic corporate headquarters. The framework of a net exposure model precludes any hedging devices, because a series of multiple exposure goals seek to maintain a zero net exposure position in each foreign subsidiary at the end of the single period. If zero net exposure can be achieved, the firm is invulnerable to any exchange rate fluctuation. A feature of the CSP, particularly attractive to firms, is that no spot exchange rate forecasts are required. The firm is viewed as being risk-averse; that is, neither willing to gain nor lose on exchange rate fluctuations relative to the domestic currency. If no exposure prevails, the firm's short-term money management profits are not subject to exchange rate risks.

In the CSP model, single-period investments and financings are permitted in each foreign country in addition to the domestic center. Cash requirements are assumed to be known for each period and for each subsidiary. The pooling concept allows transfer of funds from each country through the pool to every other country.

Let:

X_{jk} = units of foreign currency of country j denominated in US dollars transferred from country j ($j = 1, 2, 3$) to cash center pool, k;

X_{kj} = US dollars transferred from pool, k, to foreign country j;

I_j = units of foreign currency invested in country j;

I_k = US dollars invested in pool country k;

B_j = units of foreign currency borrowed in country j;

B_k = US dollars borrowed in pool country k;

O_j = deviations over net exposure goal for country j;

U_j = deviations under net exposure goal for country j;

R_j = required cash budget in country j;

S_j = maximum permissible investment in country j;

T_j = maximum permissible borrowing in country j;

i_j = percentage return on investment for duration of one month in country j;

i_k = percentage return on investment for duration of one month in pool country k;

b_j = percentage cost of financing for duration of one month in country j;

b_k = percentage cost of financing for duration of one month in pool country k.

The objective is to make the firm's net exposure in each of the foreign subsidiaries as close as possible to zero at the period's end. Using goal programming, the objective can be expressed as the minimization of the sum of the deviations over and under each net exposure goal for each country. The goal programming objective is stated as

$$\text{minimize } z = \sum_{j=1}^{3} (O_j + U_j).$$

The firm is constrained by certain operating restrictions. The first set of constraints requires the amount of funds invested at the beginning of the period plus earned interest (exposed assets) less the amount of funds borrowed at the beginning of the period plus financing expense (exposed liabilities) be equal to zero (plus or minus deviations for each country). This may be stated as

$$(1 + i_j)I_j - (1 + b_j)B_j - O_j + U_j = 0; \forall_j.$$

In other words, the constraint takes into consideration the realization that it may be impossible to obtain a zero net exposure position and allows the firm to be as close as possible to a risk-free cash position.

At the beginning of the period, funds transferred from the pool less funds transferred to the pool less investments plus financing must be equal to the required budget for the period. If excess funds exist, the budget has a negative sign. Budget constraints are

$$X_{kj} - X_{jk} - I_j + B_j = R_j; \forall_j.$$

[5] For a more detailed examination of this model, see Hollis [6].

The third constraint maintains the cash pool's balance. The total funds transferred into the pool less the total funds transferred to subsidiaries operating in foreign countries from the pool less investments plus financing in the domestic country must equal zero; or

$$\sum_{j=1}^{3} X_{jk} - \sum_{j=1}^{3} X_{kj} - I_k + B_k = 0.$$

Investments made by the foreign subsidiaries are constrained from above by establishing a maximum permissible investment level. The investment constraints are

$$I_j \leq S_j, \forall_j.$$

Financing opportunities are similarly limited in each foreign arena by

$$B_j \leq T_j, \forall_j.$$

Let the subscript n ($n = 1, 2, \ldots, 12$) represent the current application period. Subsequent to the initial application of the model ($n = 1$), the formulation must be modified to reflect implemented decision variables from the immediately preceding period, $n - 1$, and currency fluctuations. These modifications affect only the right-hand sides of the budget constraints (R_{jn}) and the cash pool balancing constraint by reflecting the fact that funds invested in the previous period plus earned interest are a source of funds for the current period. In addition, these constraints must specify the repayment of financing incremented by interest expense. All foreign transactions which are measured in US dollars must reflect previous currency value changes. This is accomplished by multiplying the amount of the transaction by the ratio of the current exchange rate (in US dollars per unit of currency j), e_{jn}, to the exchange rate at the beginning of the previous period, e_{jn-1}. The budget constraint is, thus, modified by changing the right-hand side to

$$R_{jn} - \frac{e_{jn}}{e_{jn-1}}[(1 + i_{jn-1})I_{jn-1} - (1 + b_{jn-1})B_{jn-1}]; \forall_j.$$

The necessary modification to the cash pool constraint is

$$(1 + i_{kn-1})I_{kn-1} - (1 + b_{kn-1})B_{kn-1}.$$

[6] A more detailed exposition of the DMP model is found in Hollis [7].

III. DECENTRALIZED MULTIPERIOD (DMP) MODEL

The second model, again a net exposure formulation, describes a decentralized multiperiod (DMP) money management system.[6] Each foreign-based subsidiary formulates and executes its own strategies, but must operate without the use of a centralized cash pooling center. A multiperiod decision model provides strategy decisions for the current period of one month in addition to strategies for the following two periods. Only the strategy for the current period is implemented. A moving horizon of three periods is considered at the start of each month, allowing parameters to be updated with information that would become available at the start of each month.

As in the CSP model, cash requirements for the beginning of each period are known and must be satisfied by transactions at the onset of the period. By expanding the decision horizon to several periods, more investment opportunities are considered in each country. Investments for durations of one, two, and three months may be executed. Borrowings remain at one period durations, due to the lack of published information which would have been required for model testing.

Operating within a single country, the firm faces no direct exchange rate risk. However, if the firm is a subsidiary of a corporation whose financial statements are reported in another currency, the effects of currency translations may yield substantial exchange rate currency losses or gains. Thus, any funds held in a foreign subsidiary at the termination of an accounting period are subject to translation gains or losses.

Let:

I_{ijt} = units of foreign currency of country j invested in country j's short-term money market of maturity i ($i = 1, 2, 3$) months placed in time t ($t = 1, 2, \ldots, T$);

B_{jt} = units of currency of country j obtained from short-term loans of one period in time t;

O_{jt} = deviations over zero net exposure in country j at the end of period t;

U_{jt} = deviation under zero net exposure in country j at the end of period t;

Vol. 6 1978

R_{jt} = net cash requirements in country j in time t;
S_{jt} = maximum permissible outstanding investments in country j in time t;
T_{jt} = maximum permissible borrowings in country j in time t;
i_{ijt} = percentage investment yield for duration i for purchases in country j in time t; and
b_{jt} = percentage cost of borrowing for one period in country j in time t.

For each country the model is applied with the objective of obtaining, as close as possible, zero net exposure at the end of each period. The objective function seeks to minimize the sum of the deviations from each period's net exposure goal and may be expressed as

$$\text{minimize } z = \sum_{t=1}^{T} (O_{jt} + U_{jt}).$$

The first set of constraints requires that the end-of-period net exposure be as close as possible to the prescribed goal of zero. That is, the assets maturing at the end of the period less obligations due less variations over the goal plus variations under the goal must equal zero, or

$$\sum_{i=1}^{t} (1 + i_{ij,t+1-i}) I_{ij,t+1-i} - (1 + b_{j,t-1}) B_{j,t-1} - O_{jt} + U_{jt} = 0; \forall t.$$

Budget constraints require that the cash needs of each period be satisfied. Sources of funds are single-period loans obtained within country j and maturing investments. Uses of funds include money market instruments (investments of varying maturities) or repayment of previous loans. The budget constraints are

$$B_{jt} - \sum_{i=1}^{(T+1)-t} I_{ijt} + \sum_{i=1}^{t-1} (1 + i_{ij,t-1}) I_{ij,t-1} - (1 + b_{j,t-1}) B_{j,t-1} = R_{jt}; \forall t.$$

To more fully describe the operating environment of foreign subsidiary, it is necessary to bound the maximum amount of outstanding investments and borrowings. This may be accomplished by constraints of the form

$$\sum_{i=1}^{(T+1)-t} I_{ijt} + I_{2j,t-1} + I_{3jt-1} \leq S_{jt}; \forall t, \quad \text{if } t \neq 1$$

and

$$B_{jt} \leq T_{jt}; \forall t.$$

As in the CSP model, formulations after the first application (when $n = 1$) require updating to include implemented policies that affect operating conditions in later periods. Therefore, when $n > 1$, changes are needed. First, in the net exposure goal constraints, it must be recognized that investments placed in previous periods will affect the end-of-period exposure. The right-hand side modification of the goal constraint equation is

$$\sum_{i=t+1}^{T} (1 + i_{ij1,n-i+t}) I_{ij1,n-i+t}; \forall t.$$

The budget constraints are modified by changing R_{jt} to

$$R_{jt} - \sum_{i=t}^{T} (1 + i_{ij1,n-i+t-1}) I_{ij1,n-1+t-i} + (1 + b_{j1,n-1}) B_{j1,n-1}; \forall t.$$

To maintain an upper limit on total investments, the right-hand side for the investment constraints becomes

$S_{j1} - I_{2j1,n-1} - I_{3j1,n-1} - I_{3j1,n-2}$ (when $t = 1$),
$S_{j2} - I_{3j1,n-1}$ (when $t = 2$), and
S_{j3} (when $t = 3$).

Finally, no changes are required on the maximum borrowing constraint because previously implemented financings with duration of a single period are independent of the present period's decision alternatives.

IV. NAIVE RANDOM STRATEGY (NRS)

The final money management strategy (not approximating existing practices) serves as a benchmark with which to compare the strategies obtained from the previous two models. This strategy is a naive random strategy (NRS), in which the net funds requirement for each period is determined by summing the cash requirements for each subsidiary. Next, random numbers are generated and normalised to indicate the proportion of the total funds to be obtained (invested) in each foreign country and in the domestic country. The NRS functions merely as a vehicle to facilitate the profit comparisons resulting from short-term financial practices. Any money management model performing worse than the NRS suggests the futility of that particular approach. If a naive manager can outperform another strategy by

simply generating random numbers, the time and effort consumed in a more complicated approach is egregiously wasted.

To apply this unsophisticated strategy the manager first determines the cash needs of the total firm; i.e. $\Sigma_j R_j$ (where R_j is the net cash requirement exogenously determined for country j at the beginning of the period). If the net figure is positive, financing must be obtained; if negative, investments are placed. If the net figure is zero, no action is taken. After obtaining the net cash position, four random numbers are generated. These numbers are normalized to obtain weighting factors, w_j, which when multiplied by total cash needs indicate the amounts of US dollar equivalents to be invested, I_j, or borrowed, B_j, from each of the j countries (where $j = 1, 2, 3, 4$).

Subsequent to the $n = 1$ (first application period) case, certain modifications are required to ensure repayment of loans or inclusion of matured investments and to incorporate currency value changes measured in US dollars. That is, the net cash needs for the MNC are found by summing the following:

$$\sum_{j=1}^{4}\left[R_{jn} - \frac{e_{jn}}{e_{jn-1}}(1 + i_{jn-1})I_{jn-1} + \frac{e_{jn}}{e_{jn-1}}(1 + b_{jn-1})B_{jn-1}\right].$$

where:

R_{jn} = net cash needs for country j in period n;
I_{jn-1} = investments placed in the previous period, $n - 1$, in country j;
B_{jn-1} = financing obtained in country j in $n - 1$;
i_{jn-1} = percentage yield from investment in country j placed in period $n - 1$;
b_{jn-1} = percentage cost of financing in country j obtained in period $n - 1$;
e_{jn} = exchange rate in US dollars/unit of currency of country j at the beginning of the current period, n;
e_{jn-1} = exchange rate in US dollars/unit of currency of country j at the beginning of the previous period, $n - 1$.

The coefficients required for updating the NRS are known with certainty. That is, the manager knows borrowing costs or investment yields, and the exchange rate for the previous period. Further, the current exchange rate, e_{jn}, is known. In the case of domestic transactions, both e_{jn} and e_{jn-1} have values of one (illustrating that US investments and borrowings are not subject to exchange rate fluctuations).

V. EXPERIMENTAL DESIGN AND PROFIT RESULTS

This section compares the results that would have been experienced had NRS, CSP and DMP money management strategies been implemented each month during the period September 1974 through August 1975. For each strategy, the associated model (as described in preceding sections) was used to obtain recommended monthly transactions. The transactions were then submitted to a simulator that determined what would have happened had an MNC consistently followed a strategy for twelve consecutive months. It is important to notice that results presented do not represent profits that are predicted by models; rather, the results show the profits that would have been experienced had a decision maker implemented strategies suggested by models.[7]

The reported profit performances of alternative strategies are before-tax consolidated profits measured in US dollars. Profits are comprised of gains from short-term investments reduced by costs of short-term financings adjusted for exchange rate changes on net exposed assets.

A randomly determined set of cash requirements is established for each foreign subsidiary in each of the twelve months (see Appendix). Cash needs for each period range from −$25,000 to $25,000. Negative and positive signs represent, respectively, cash-rich and cash-poor subsidiaries.

In all models, foreign financing is limited to a maximum of $50,000 obtained in each country for a period of thirty days.[8] Domestic financing is available only in the centralized models (CSP) with no limits. The short-term rate to prime borrowers is obtained from the *Bimonthly FFO Updater and Forecaster of Interest Rates* [1].

For all models, foreign investments are constrained to a maximum of $150,000 per period per country, while no limit is placed on domes-

[7] Exchange controls, of course, could prevent the firm from implementing desired transactions.

[8] No multiple period financing is permitted due to lack of published information.

tic investment. In the multiperiod model (DMP) investment constraints are refined to specify a maximum of $150,000 outstanding per period.[9] Thus, if the maximum investment is placed with a more-than-one period maturity, no other investments may be executed until the investment matures.

Short-term investments selected for inclusion are similar with respect to risk and liquidity characteristics of the US Treasury Discount Bill. Yields for the US bills are obtained from *Moody's Bond Record* [12]; United Kingdom rates are obtained from the *Financial Times* [3]; and the Canadian rates are obtained from *International Financial Statistics* [9].[10] The *Monthly Report of the Deutsche Bundesbank* [11] is the source for German call money rates.

Monthly before-tax money management profits are calculated from the application of each model on a cash basis for the hypothetical MNC. Profits are defined to be before-tax consolidated gains on investments resulting from interest and revaluations from parity changes (or reductions from devaluation) less financing charges and revaluation changes (or increases from devaluation). Profit results are detailed in Table 1.

An inferential test is presented to afford generalizations of the research results. The hypothesis is that the respective applications of the three money management strategies result in no significant differences among monthly before-tax consolidated profits for the period September 1974 to August 1975. The hypothesis is tested using the nonparametric Friedman test for analysis of variance for dependent groups. For each month, the profits resulting from the three strategies are ranked and the ranks are summed. If the null hypothesis is true, the sum of the ranks used should be equivalent for each model. The sum of the ranks used in the Friedman test are, respectively, 26 (NRS), 22 (CMP), and 24 (DMP). The value of the calculated Friedman statistic is 0.7 which would require at least a 0.975 significance level for rejection according to tables found in Lehmann [10]. At generally accepted levels of significance, then, there are no significant differences in the profit levels of the three models.

TABLE 1. MONTHLY CASH PROFITS (IN US DOLLARS)

Month	Strategy NRS	CSP	DMP
9/74	341.22	274.69	0.00
10/74	−441.03	−244.84	−302.95
11/74	−66.82	−27.53	−10.56
12/74	267.55	146.88	35.68
1/75	−46.60	210.36	229.17
2/75	559.96	144.32	507.13
3/75	−164.77	41.92	767.06
4/75	−30.35	−69.13	167.91
5/75	−54.86	−130.50	−254.90
6/75	29.64	−47.95	53.82
7/75	424.34	−105.27	−442.64
8/75	−83.03	−425.61	−1,683.62
Total	735.25	−232.68	−933.90

VI. CONCLUSIONS

The inferential conclusions reached in this research study are only generalizable to the MNC operating in the countries specified for the sampled period for any series of cash requirements given the set of assumptions invoked. The net exposure models performed no better than a random "dart throwing" approach. This is not surprising in the volatile period of fluctuating exchange rates. Most of the pure net exposure models were developed and reported prior to February 1973 when the flexible exchange system was introduced for most major currencies.

Even the extension of a net exposure model into a multiperiod formulation does not yield a viable foreign exchange risk management model. The multiple-period formulation introduces matching exposure problems. Equal dollar assets and liabilities obtained in different periods do not yield net exposure of zero (when considering translation procedures for independently managed subsidiaries) if incremental movements in exchange rates differ. Projected zero net exposure in a currency (measured in the foreign currency with different transaction dates) does not leave the MNC invulnerable to exchange rate fluctuations. Indeed, the naive random approach results in no significant profit differences. Firms currently operating with

[9] The parameters of this research have been fixed at particular values to prevent unlimited recourse to money markets.

[10] Given the unavailability of the term interest structure of short-term Candian bill rates, it was assumed that one-month rates were 31% and that two-month rates were 63% of the three-month rates which were readily available.

simple net exposure schemes should re-evaluate their management procedures.

Strategies resulting from the multitemporal model, DMP, have strong implications for the importance of the length of planning horizons. The closer to time of implementing transactions, the more accurate future strategies approximated the actual transactions. That is, the formulation of planning is better with more current information.[11] Use of the three-period time horizon (while not rigorously tested in this research) may be better than a longer number of periods. With longer time horizons, additional inaccuracies (forecasts on future parameters) are introduced as well as additional computational burdens incurred.

International short-term money management affords many avenues for continued research. Forecasting methods for future spot exchange rates, investment rates, financing rates, and forward contracts need to be explored and validated. Strategies which include all exposed assets and liabilities rather than monetary classifications would be invaluable to the MNC.

Sensitivity analysis of the models to the robustness of parameters' values (which were assumed to be fixed at certain levels in this research) needs to be explored. More specifically, simulation methods can be employed to ascertain the effects of profits by varying the range of the randomly determined cash requirements and levels of maximum permissible investments and financing.

In conclusion, MNCs which are currently making decisons based on simple zero net exposure methods should re-evaluate their practices. It appears that some degree of exposure may be desirable.

ACKNOWLEDGEMENTS

The author wishes to thank Dale VorderLandwehr, Roger Eck, and Joe McGuire for their helpful comments on this research.

[11] The author wishes to thank the referees for pointing out that there may be a case for time-dependent weightings for DMP deviational goals. For example, the more recent the period, the more costly should be weights employed in the objective function for exposure deviations since only the current period's solution is implemented.

REFERENCES

1. *Bimonthly FFO Updater and Forecaster of Interest Rates.* Various Issues. Business International Corporation, New York.
2. Eiteman DK & Stonehill AI (1973) *Multinat. Bus. Fin.* Addison–Wesley, Reading, Massachusetts, US.
3. *Financial Times.* Various Issues. London.
4. Giddy IH & Dufey G (1975) The Random Behavior of Flexible Exchange Rates: Implications for Forecasting. *J Int. Bus. Stud.* **6** 1–32.
5. Goeltz GK (1972) Managing Liquid Funds Internationally. *Columbia J Wld. Bus.* **7** 59–65.
6. Hollis MS (1976) Illustration of the Behavior of a Single-Period Net Exposure Model. Working paper. University of California, Irvine, US.
7. Hollis MS (1976) A Mathematical Multiperiod Decentralized Net Exposure Model. Working paper. University of California, Irvine, US.
8. *International Financial Statistics.* Various Issues. International Monetary Fund, Washington.
9. Korth C (1971) Survival Despite Devaluation: Reducing Overseas Exposure. *Bus. Horiz.* **14** 47–52.
10. Lehmann EL (1975) *Nonparametrics: Statistical Methods Based on Ranks.* Holden–Day, San Francisco.
11. *Monthly Report of the Deutsche Bundesbank.* Various Issues. Deutsche Bundesbank, Frankfurt, Germany.
12. *Moody's Bond Report.* Various Issues. Moody's, New York.
13. Teck A (1970) Financial Planning in Nations with Fluctuating Currencies. *Mgmt. Acct.* **58** 25–28.
14. Teck A (1971) Control Your Exposure to Foreign Exchange. *Harv. Bus. Rev.* **52** 66–75.

ADDRESS FOR CORRESPONDENCE: *Ms Martha Hollis, Assistant Professor of Management Science, Graduate School of Administration, University of California, Irvine, CA 92717, USA.*

APPENDIX

RANDOMLY GENERATED SUBSIDIARY MONTHLY CASH REQUIREMENTS* (US DOLLARS)

	Subsidiary		
Month	United Kingdom	Germany	Canada
9/74	−10,097	−10,480	−14,194
10/74	12,831	22,368	24,830
11/74	20,632	−24,130	−18,738
12/74	−20,247	10,365	−17,617
1/75	−11,087	−14,158	13,602
2/75	10,600	−15,290	15,398
3/75	14,259	15,404	−14,778
4/75	13,452	−14,542	23,495
5/75	−20,662	12,595	18,059
6/75	−19,008	20,665	−15,053
7/75	18,062	−20,335	12,566
8/75	20,474	10,526	17,983

* Negative values represent excess cash.

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 257–266
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0257$02.00/0

Evaluating Time/Cost Factors of Implementation via GERT Simulation

BERNARD W TAYLOR III

Virginia Polytechnic Institute and State University

K ROSCOE DAVIS

University of Georgia

(*Received June 1977; in revised form November 1977*)

Managers in striving for system implementation often disregard the time and cost that must be expended in achieving implementation success. This paper explores the use of GERT network analysis as a tool for planning and ascertaining the expected time and cost of implementation. The supporting base for the study is an implementation process that employs computer gaming for achieving organizational change.

INTRODUCTION

TWO OFTEN overlooked factors related to the development and implementation of management systems are the cost and time involved in implementation. Because of potential failure during implementation [1, 2, 4, 5, 7], managers often focus attention on achieving implementation with little regard for the price of success. The purpose of this paper is to describe a generalized GERT simulation model of an implementation process. The model provides statistical data on the time and cost of implementation. Such a model enables the planning of the implementation process so that excessive disruptions in the work environment (i.e., budget, scheduling, etc.) can be avoided. To describe the GERT model, a definitive implementation process is required. The implementation process employed in this paper is the simulation–gaming process developed by Davis and Taylor [3].

THE IMPLEMENTATION PROCESS

Before describing the GERT model, the Davis–Taylor [3] process for implementation will be reviewed. This implementation process is a methodology for achieving the three-step change process (i.e. unfreezing, changing and refreezing) which comprises the Lewin-Schein theory of change [6, 8, 9]. Through empirical testing, Sorensen and Zand [10, 12] found that a high level of unfreezing, changing, and refreezing results in a high level of success in implementation, while low levels of the change steps result in low levels of implementation success. However, while Sorensen and Zand demonstrated that the change theory contributed to the success of implementation, the authors failed to identify a methodology for achieving the Lewin–Schein cycle. Specifically, a "change agent" which accomplished the change steps was not identified.

The Davis–Taylor model [3] fills this gap by defining a methodology which acts as the change agent. In the Davis–Taylor model a computerized simulation game is employed as the change agent. The gaming approach is an on-line computer system process which compliments the steps of the Lewin–Schein theory: (1) unfreezing via game/system introduction, (2) changing via game playing, and (3) refreezing via "on-line" deployment of the system.

A frame work of the gaming methodology and its relationship to the three-stage Lewin–Schein change process is depicted in Fig. 1. The first component of the model, establishing sys-

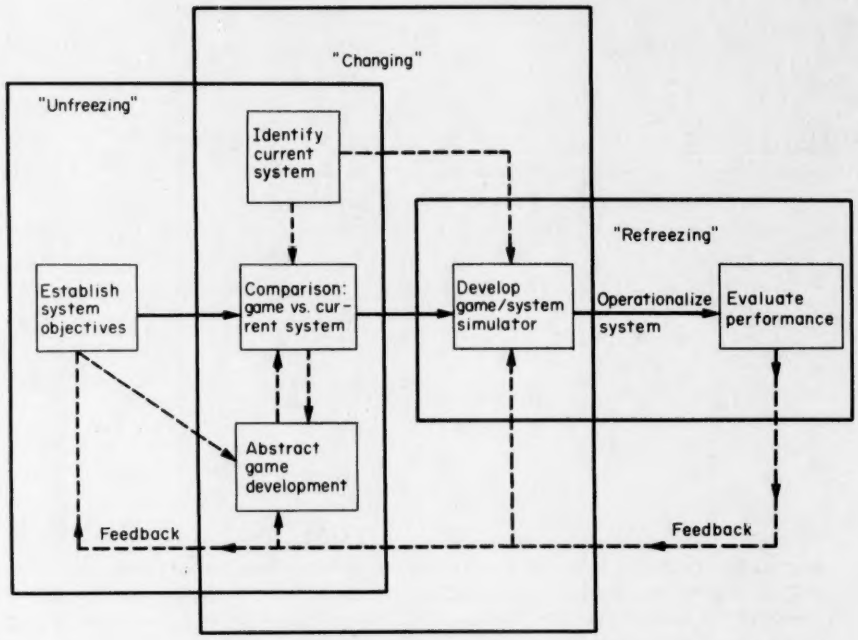


FIG. 1. *The implementation process.*

tem objectives, initiates the unfreezing process by establishing high performance levels for managers. This creates uneasiness and dissatisfaction for the manager. The unfreezing stage is completed by identifying the current system and comparing it with an abstract game of the system to be implemented.

The structure of the gaming activity includes game play, followed by critique sessions where the systems analyst and/or planners and managers interact. The gaming activity results in learning for both parties—the manager learns the capabilities and features of the new system, while the analyst/planner ascertains the needs of the manager and refines and develops the new system. This learning process serves to complete the unfreezing process and begin the "changing" stage.

Once the abstract game playing is completed and all inputs are assimilated, a simulation/game model of the actual system is developed. This system-model is then employed as a game. While playing the new game, the manager actually learns the system to be operationalized. This completes the change step and begins the refreezing step.

Refreezing is continued by demonstrating the system's effectiveness in achieving results and by operationalizing the system. The transition from game playing to actual implementation should be smooth if prior learning experiences have been achieved. The refreezing step is completed if the system is evaluated and found to achieve original objectives. If the system is not functioning as expected discrepancies must be corrected or the manager will discard the new system.

The game-implementation methodology has been found to greatly enhance the probability of successful implementation [3]. However, such a methodology consumes resources, i.e. time and money. The process of game development, game playing, and operationalizing may encompass an extended period of time, which in retrospect may be undesirable or unwarranted by the firm. Time lags during implementation can result in decreased output and productivity, lost sales, and disrupted production schedules. Similarly, implementation costs may become extensive and may make implementation inappropriate when compared with cost savings expected from the new system. In the extreme case, the cost of successful implementation may be greater than the accumulated savings garnered by the new system.

Because of the possibility of cost and time overruns, some form of planning must be undertaken prior to implementation to ascertain

if the system warrants implementation. If implementation is to be accomplished, appropriate resources (time and dollars) must be made available. Since the implementation process is a step-by-step procedure (i.e. network), GERT network analysis is an appropriate tool for simulating the implementation process. The GERT simulation model will yield statistical data concerning expected time and cost of system development and implementation. (For sources on GERT analysis, see [11].)

A GERT NETWORK OF THE IMPLEMENTATION PROCESS

A case example computerized inventory control system will be employed to describe and demonstrate the use of GERT for planning the implementation process. The system, developed for use in a large assembly manufacturing plant, was designed to provide decisions for maintaining proper inventories, minimizing inventory costs, and preventing unwarranted backlogs. Limited discussion of the system will be provided; rather the GERT model for planning its implementation will be presented.

The GERT model

The inventory control system examined in the study had been developed a number of months prior to the time that implementation by use of the Davis–Taylor process was considered. Problems had been encountered when the system was initially operationalized. Line managers argued that the system forced them to operate their lines with considerably less inventories than were "operationally feasible". Marketing managers argued that they were losing sales because of stock-outs. Top management was of the opinion that the system was theoretically sound but at least two problems existed: (1) managers more than likely didn't understand the operations of the system, even though formal seminar-implementation classes had been conducted; and (2) the system forced managers to reduce inventories to theoretical minimums; however, analysts assured top management that slight changes could be made in the system such that individual managers could control the rate of reduction and thus avoid drastic inventory reductions that could result in missed customer deliveries.

The decision was made by top management to "drop the system" if the problems could not be resolved and if the additional costs of implementation could not be cost justified. The Davis–Taylor implementation process was proposed for implementing the system, since it appeared to provide a means for "training" and a means for "system modifications," if needed. The key questions raised were, "What would be the cost of employing the process; what was the probability that the results would be the same as when the system was first implemented; and finally, what was the expected length of time for achieving full implementation."

Since GERT network analysis could provide statistical data related to the questions raised (see [11]), it was adopted as a suitable vehicle for studying the proposed implementation approach. The gaming-implementation framework depicted in Fig. 1 was transformed into the GERT network shown in Fig. 2. Table 1 is a complete description of all activities, nodes, branch probabilities, and time and cost distributions for the inventory system.

The GERT network closely resembles the gaming framework of Fig. 1. Activities such as establishing system objectives, identifying the current system, abstract game development, and game playing are included. Note from Table 1 that the time distributions for these, as well as other activities is a Beta probability function. Since time estimates are typically subjective in nature, the Beta distribution is appropriate (even though GERT has the capabilities for simulating numerous other distributions). Each time distribution for the inventory system was defined by three subjective estimates (optimistic, pessimistic, and most likely). Fixed and variable costs were also subjectively determined for each activity. The fixed cost is realized when the activity is undertaken, while variable cost is determined by the duration of each activity.

For the inventory system, start up of the implementation process, indicated by activity 10–11, consisted of identifying the personnel and initial resources (space and equipment) necessary to begin the implementation process. The personnel group identified consisted of a systems analyst, a management scientist and a computer support specialist. A fixed cost of $500 was assigned to the activity.

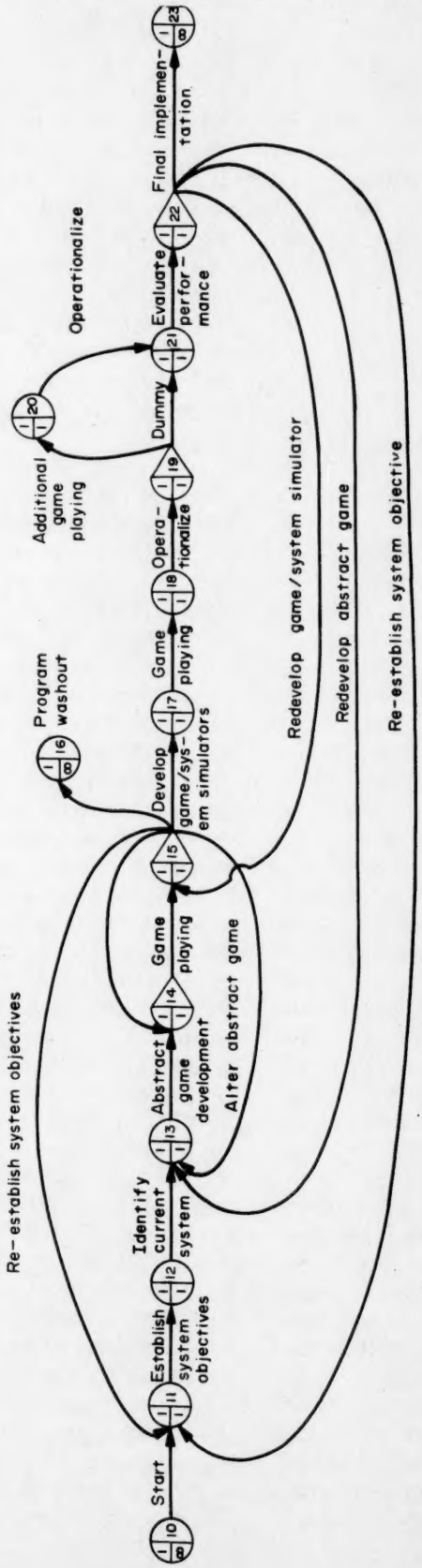


FIG. 2. *The GERT implementation process.*

TABLE 1. ACTIVITY DESCRIPTIONS WITH TIME AND COST ESTIMATES

Activity (nodes)	Activity description	Probability of occurrence	Time estimates (days)			Distribution	Cost estimates ($)	
			Min	Mode	Max		Set-up	Variable
10–11	Start process	1.00		0		Constant	500	0
11–12	Establish system objectives	1.00	10	20	30	Beta	0	100
12–13	Identify current system	1.00	5	10	20	Beta	0	75
13–14	Abstract game development	1.00	15	30	60	Beta	2500	100
14–15	Game playing	1.00	20	35	65	Beta	0	150
15–16	System washout	0.05		0		Constant	0	0
15–17	Develop game simulator	0.60	15	30	60	Beta	3000	100
15–11	Reestablish objectives	0.01		0		Constant	0	0
15–13	Alter abstract game	0.03		0		Constant	0	0
15–14	Participate in more gaming	0.31		0		Constant	0	0
17–18	Game playing	1.00	20	40	65	Beta	0	125
18–19	Operationalize	1.00	10	25	50	Beta	1000	100
19–20	More game playing	0.20	5	15	30	Beta	0	125
19–21	Dummy	0.80		0		Constant	0	0
21–22	Evaluate performance	1.00	7	15	30	Beta	0	75
20–21	Operationalize	1.00	8	15	30	Beta	0	100
22–23	Final completion	0.90	5	10	25	Beta	0	100
22–15	Redevelop simulator	0.07		0		Constant	0	0
22–13	Redevelop game	0.02		0		Constant	0	0
22–11	Reestablish objectives	0.01		0		Constant	0	0

Activity 11–12 represents the first major phase of the implementation process. For implementing the inventory system under the Davis–Taylor process this required assessing the time and cost that would be required to meet with top management to ascertain objectives and then disseminate the objectives to production and marketing managers who would use the system. The activity would involve working with each middle manager to identify the basis for the objectives of the proposed system. The activity was estimated to take between ten and thirty days at a cost of $100 per day.

Activity 12–13, "Identify Current System," involved assessing the time and costs involved in working with middle managers to identify problems and weaknesses of the existing system and comparing the existing system output with management's established objectives. In terms of the Davis–Taylor process this activity would complete the unfreezing stage and initiate the change phase of the process. A time estimate of five to twenty days (with a most likely estimate of ten days) was established as the time that would be required to complete the activity; $75 per day was estimated to be the cost associated with the activity.

Development of a computerized abstract game is the activity involved at node 13–14. This is the stage of the implementation process where top management objectives are incorporated into an abstract computerized model (game) of the "to be developed/modified system." It was estimated that it would require between fifteen and sixty days to develop the abstract model, with an initial cost of $2500 and a variable cost of $100 per day. The $2500 set-up cost was the projected cost of capital equipment (on-line terminal) necessary to support development of the model. Variable costs would be programming costs, cost of developing questionnaires, and costs of developing training session transparencies.

Activity 14–15 is the point at which the abstract model is employed as a game. The game activity acts as a catalyst to complete the unfreezing phase of the implementation process—managers, through game playing, fully comprehend the defined objectives. The game playing involves managers making decisions in game playing session via an on-line

terminal. Results of game play are discussed among the implementation staff and managers. It was estimated that this activity would consume between twenty and sixty-five days at a cost of $150 per day.

After the abstract game has been played for a reasonable period of time, decisions must be made at node 15 about the perceived future success of implementing the inventory system. If the entire approach appears to be futile and failure is eminent, the project is washed out (activity 15–16). This reduces the cost of completing the entire project and incurring system failure. Alternatively, at node 15 it may be decided that some activities should be repeated if implementation has not proceeded as expected. These alternatives are denoted by activities 15–14, 15–11, and 15–13. In each case reworking will result in a substantial increase in time and costs overall implementation process. In the GERT planning model the estimation of the node probabilities associated with the different alternatives is quite difficult since historical data is usually nonexistent or limited. Since the model is quite sensitive to these probabilities care should be exercised in the estimation process. Note in Table 1 that the sum of probabilites of all activities leading from node 15 is equal to one.

Unless it is necessary to alter the abstract game (node 15–13), participate in more gaming (node 15–14), reestablish objectives (node 15–11), or the system washes-out (node 15–16), the next activity in the implementation process is development of a game simulator that mirrors the actual inventory system (node 15–17). This game is developed in the same manner as the abstract game but is in reality a simulation of the actual system. The time for this activity was estimated to require between fifteen and sixty days at a fixed cost of $3000 and variable costs of $100 per day.

After the game/simulator has been developed, it is played (activity 17–18) in the same manner as the abstract game (i.e. via game sessions with production and marketing managers). It was estimated that this would be the longest activity in the implementation process since this is where managers experiment and learn the actual system to be operationalized. This activity is extremely important since it completes the change step and initiates the refreezing step in the implementation process.

Following the learning of the inventory system via game playing, the new system is operationalized (activity 18–19). Operationalizing the system requires modifying the simulator to reflect any problems or constraints noted during the gaming sessions. Ten to fifty days, with an average of twenty-five days, was the time estimated to be required to operationalize. A set-up cost of $1000 was estimated for additional capital equipment.

After the system has been operationalized it may be determined that the system has not been completely "learned"; additional game play may be required (activity 19–20)—this problem should result only for a small portion of the managers. Following the operationalizing of the system a period of time should be set aside to evaluate the performance of the system (activity 21–22). For the inventory system, a time estimate of between seven and thirty days was established for system evaluation.

If the system, as a result of evaluation, is found to be successful (i.e., the objectives are accomplished), the system can be declared complete (activity 22–23). However, if it becomes apparent during system evaluation that the system is not meeting the established objectives some or all prior stages of the implementation process must be repeated. These activities are denoted in Table 1 by activities 22–15, 22–13, and 22–11. As was the case at node 15, probabilities of occurrence must be estimated for each activity leading from node 22. Note that in the case of node 22, however, the probabilities of rework should be smaller than those that occurred at node 15. For the inventory system it was estimated that there was a ninety percent chance that rework would not be required. Again, as was the case at node 15, estimating the node probabilities is important in obtaining realistic cost and time results for the entire implementation process.

Simulation results

The implementation GERT network was iterated 1000 times in order to obtain operating statistics for time and cost. Table 2 gives the summary of time statistics for the project, while Table 3 gives the summary of cost statistics. Data in these tables show that the implementation process would on the average

TABLE 2. SUMMARY OF TIME STATISTICS FOR IMPLEMENTATION PROCESS

	Time (days)				
Event	Prob.	E(t)	σ_t	Min. t	Max. t
Process completion	0.914	275	73	184	679
Process washout	0.086	131	56	78	319

TABLE 3. SUMMARY OF COST STATISTICS FOR IMPLEMENTATION PROCESS

	Cost ($000's)				
Cost	Prob.	E(c)	σ_c	Min. c	Max. c
Process cost	0.914	37	10	26	92
Washout cost	0.086	17	8	10	45

require 275 days for full implementation, with an expected cost of $37,000.[1]

However, recall that top management was interested in more than the expected cost and time for implementation. The question had been asked, "What is the probability that the implementation process will result in failure, as was the case when the system was first implemented." Tables 2 and 3 provide data on this question. There is an 8.6 percent chance of implementation failure at an expected cost of $17,000. Note that the expected cost of failure is almost half the expected cost of success. This results because a lengthy period of time elapses (on the average 131 days) before the probability of failure is evaluated. Even though there was the possibility of large losses if failure occurred, the decision was made to employ the processes.

Operation results

The Davis–Taylor process was employed to reimplement the inventory system. Thirty-seven line managers were involved in the implementation process. The system was successfully implemented. Two hundred and eighteen days were required to implement the system in twenty-one different manufacturing operations. Rework was required at two activities in the process. At the operationalizing activity it was found that line workers who had not been involved in the early parts of the implementation process did not fully comprehend the system's operations. Game playing exercises, using the game/simulator (activity 19–20), were employed to familiarize the workers with the system. During the evaluation of the system (activity 21–22) it was determined that additional flexibility was needed in the system; the simulator was modified (activity 21–15) to reflect the desired changes. Mini-game playing sessions were conducted to reflect the changes in the system. The changes were then incorporated into the final system. Total costs of the implementation process was $38,200.

Problem areas

One of the key problem areas related to employing GERT to study the implementation process is development of probability, time, and cost estimates (Table 1). Fortunately, the Davis–Taylor implementation process had been employed in implementing another system prior to being considered for use in implementing the inventory control system. Rudimentary time and cost data thus existed from which estimates could be made.

The prior system in which the Davis–Taylor process was employed was a resource-allocation-optimization system; this system differed significantly from the inventory control system. However, since the steps of the implementation process were the same under both systems, prior time and cost data from the resource system were valid bases for making time and cost estimates for the inventory control system. Data from the resource allocation were modified through interactions with managers, analyst, and system operators.

[1] Figures 3 and 4 are frequency histograms of the time and cost statistics of Tables 2 and 3.

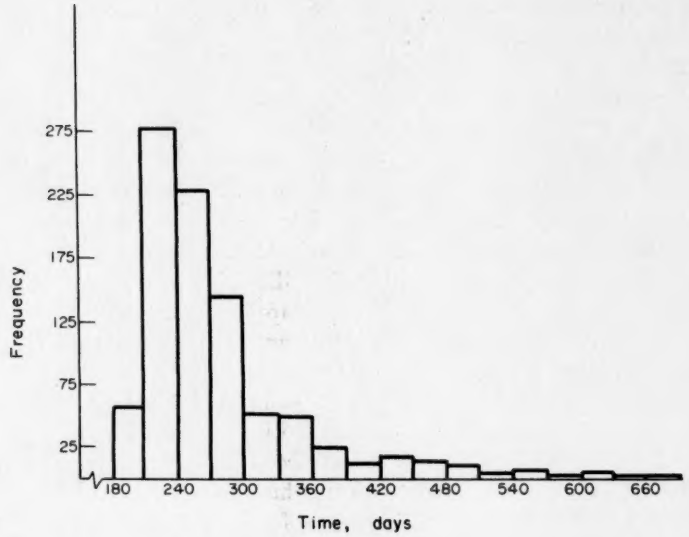


FIG. 3. *Time to completion of implementation process.*

The most useful data were prior time requirements. A large volume of data existed on *time required to develop the abstract game, initial game playing time, simulator development time, time employed in the second game playing exercises,* and the *time required in operationalizing the system.* A limited amount of data existed on start up activities such as *time required in establishing objectives* and *time required to identify the existing system.* Some differences of opinion existed as to the expected times that would be required to perform the implementation steps for the inventory system. The extreme estimates in each case were used as the pessimistic and optimistic values.

Cost estimates were developed for each of the implementation steps by relying on data from the resource allocation system. However, the only useful cost data were the fixed or rather "set-up" cost. Variable cost data for the inventory control system were primarily the cost associated with management, analyst or

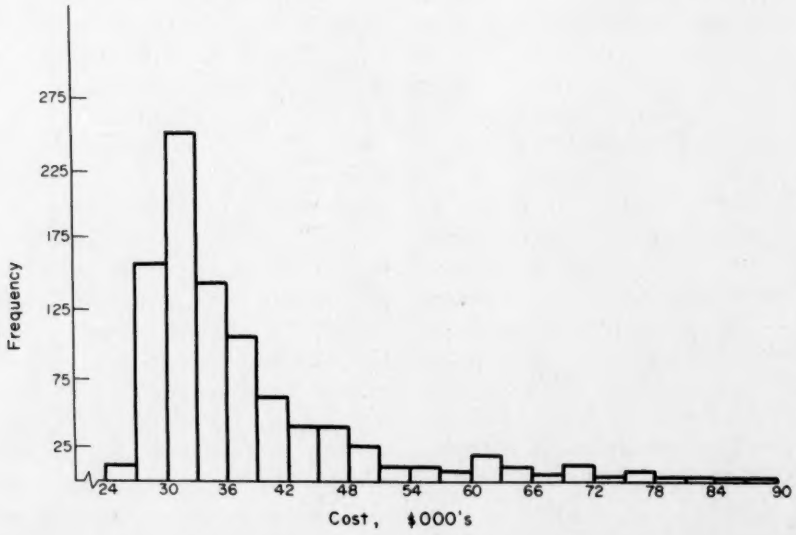


FIG. 4. *Cost of completion of implementation process.*

user (operator) involvement. These costs were computed by using man-hour costs taken from cost accounting records and by estimating total man-hour involvement. To make the cost estimates compatible with the time estimates (the total number of days for completion of an activity) it was necessary to convert total man-hour costs to cost per man-day. For example, for the game playing activity (node 14–15) it was estimated that thirty-one line managers would be involved in game sessions that required approximately 4.5 hours per manager. Cost accounting records indicated the average line manager's rate per hour was $37.50. Total cost of the activity thus was estimated to be $5,250.00. The prior time estimate for completing the game playing activity was 35 days (5 game sessions were to be conducted with 6–7 managers per session; the sessions were to be conducted on the average, every 7 days). The conversion of total activity cost to average cost per man-day resulted in variable costs that on observation appeared somewhat low. The GERT simulator required compatability of units, thus the conversions were necessary.

Probability estimates for the inventory system, although subjectively determined, were based upon the results from the resource-allocation system. The only activity where managers' estimated probabilities differed significantly was node 22–23 (final completion). Most managers were of the opinion that the system would meet performance specification (final completion) without having to recycle prior steps in the implementation process. Some managers, however, expressed doubt as to this resulting. The actual probability employed in the model was the average of the expressed probability estimates.

CONCLUSIONS

The primary use of GERT simulation in this study has been to aid the planning function associated with developing (redeveloping) and implementing a system. Often managers become overly concerned with "getting the system on-line", and ignore the time and costs which may be required in developing and implementing the system. The data provided by the GERT model can aid management in evaluating the development-implementation process prior to beginning development of the system. Production schedules, market strategies, and budget decisions can be coordinated to fit the development-implementation process.

Through the paper the Davis–Taylor process was used and emphasized as the process or procedure suitable for implementation. The GERT network evaluation process is not limited to this particular implementation process. GERT is a general framework which can be adapted to any particular development/design/implementation framework. Activities which may be peculiar to a particular system or design/implementation can easily be incorporated into a GERT model.

A secondary benefit that can accrue from using the GERT evaluation process is enhancement of understanding the requirements of implementation. The evaluation process forces management to identify the activities that will be used in developing and implementing a system. It also forces management to place time, cost, and probabilities on different activities, a factor that is generally ignored in developing most systems.

The GERT evaluation process is not a panacea for the often ignored questions of implementation time and cost, but it can provide some answers.

REFERENCES

1. Ackoff RL (1960) Unsuccessful case studies and why. *Op. Res.* **8**(2), 259–263.
2. Claycamp H (1971) Letter from the Chairman. TIMS College on Marketing, Newsletter **5**(3).
3. Davis KR & Taylor BW (1976) Addressing the implementation problem: A gaming approach. *Decis. Sci.* **7**(4), 677–687.
4. Dyckman TR (1967) Management implementation of scientific research: An attitudinal study. *Mgmt Sci.* **13**(10), 612–620.
5. Harvey A (1970) Factors making for implementation success and failure. *Mgmt Sci.* **16**, B312–B320.
6. Lewin K (1947) Group decision and social change. In *Readings in Social Psychology* (Eds Maccoby, Newcomb and Hastley) pp. 197–211. Holt, Rinehart and Winston, New York.
7. Malcolm DG (1965) On the need for improvement in the implementation of OR. *Mgmt Sci.* **11**, B48–B58.
8. Schein EH (1964) The mechanism of change. In *Interpersonal Dynamics* (Eds Bennis, Schein, Steele & Berlew) pp. 362–378. Dorsey Press, Homewood IL.
9. Schein EH (1969) *Process Consultation: Its Role in Organization Development.* Addison-Wesley, Reading, Massachusetts.
10. Sorenson RE & Zand DE (1974) *Overcoming Problems of Decision Sciences Implementation.* Paper presented at 1974 Southwestern AIDS, Dallas, Texas.

11. WHITEHOUSE GE (1973) *Systems Analysis and Design Using Network Techniques.* Prentice–Hall, Englewood Cliffs, New Jersey.
12. ZAND DE & SORENSON RE (1975) Theory and change and the effective use of management science. *Admin. Sci. Q.* **20**, 532–545.

ADDRESS FOR CORRESPONDENCE: *K Roscoe Davis, Esq, Associate Professor, Department of Risk Management, Insurance and Management Science, College of Business Administration, The University of Georgia, Athens, Georgia* 30602, *USA.*

Vol.
197

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 267–272
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0267$02.00/0

A New Model for Aggregate Output Planning

THOMAS J HINDELANG

JOHN L HILL

Drexel University, Philadelphia

(*Received March 1976; in revised form January 1978*)

Aggregate production planning has been aided in the recent past by the advent of mathematical programming models and the improvement in computer solution techniques. Such models have only considered single objectives; this paper formulates a multi-objective production planning model as a goal program and discusses the problems involved in its implementation.

Vol. 6
1978

INTRODUCTION

THE FOCUS of this paper is to formulate an aggregate output planning approach which capitalizes on the strengths of goal programming (GP) in incorporating multiple behavioral and economic considerations into the analysis. The new model will be formulated for a single plant with several departments each having the capability to produce one or more products. The model allows readily for disaggregation in several ways. The decision variables incorporated in the model are the aggregate production rate, the work force level, the aggregate amount of overtime worked, and the total amount of subcontracting. These four controllable factors are the major strategies available to management which enable the adjustment of inventory levels so as to smooth out demand fluctuations and achieve an hierarchy of relevant goals.

The next section of the paper surveys the aggregate output planning problem and relevant research to date. The new formulation is then developed, and finally, implementation of the model is discussed.

STATEMENT OF THE PROBLEM

Aggregate output planning (AOP) is the essence of a manufacturing firm's intermediate production planning phase, having a one to a twelve-month planning horizon wherein decisions are made relative to the production rate, workforce level, capacity utilization, amount of subcontracting to be undertaken, inventory levels to be maintained, and the procurement of raw materials. AOP concentrates on determining which combination of these decision variables should be utilized in order to optimally adjust to demand fluctuations within the constraints imposed by long-range plans. Without question, the economic, goal-oriented, and behavioral significance of the AOP problem requires careful balance and coordination between robust models and managerial talent in all of the firm's sub-systems which contribute to the decision process. Smooth integration of these subsystems is required in order to achieve effective and efficient performance of the overall firm as it interacts with the environment.

Research to date in the AOP area has been rather extensive; surveys can be found in Groff and Muth [3, Chapter 13] and Buffa and Taubert [1, Chapters 5–7]. To date, two models have been suggested which apply GP to the AOP problem setting. The former was by Jaaskelainen [4] which merely has four goals: one for meeting demand requirements and one for each of the three major costs (stockout cost, costs associated with changing employment levels, and inventory costs) which are relevant in AOP. He does not incorporate any of the

other relevant dimensions of the problem such as: interactions among various departments and sub-systems of the firm; management policies or goals relative to stockouts, frequency of changes in employment levels, and the public image of the firm; and worker motivation, satisfaction, and performance. The more recent GP model was proposed by Goodman [2] who utilized GP as a method of linearizing the quadratic cost terms of the AOP problem and considering higher order cost terms. His formulation also ignores the relevant dimensions of the problem just cited and also does not make use of GP's ability to penalize different cost components on various priority levels in the objective function.

The new formulation recommended below extends the above GP models and capitalizes on the following aspects ignored by them. First, the overall cost function is segregated into its three major components—work-force size, production rate, and inventory costs—so that management has additional flexibility in penalizing deviations from the various types of costs depending on the firm's characteristics (e.g., its size, industry, level of mechanization, stability of demand, cost of products, financial viability, etc.) and management's perceptions of tradeoffs among the cost components. Second, behavioral objectives, managerial insights and policy considerations are incorporated. Third, the model optimizes the aggregate production variables as well as determining the optimal product mix given the multidimensional problem setting described above. We now turn to a discussion of the new formulation.

THE GOAL PROGRAMMING FORMULATION

The following specific goals are incorporated in the new formulation:

1. To come as close as possible to a specified worker productivity goal;

2. To promote worker motivation through job rotation and labor force stability;

3. To limit the dollar cost associated with work force level to a specified amount or to minimize this cost;

4. To limit the dollar cost associated with production rate and subcontracting to a specified amount or to minimize this cost;

5. To limit the dollar cost associated with inventory levels to a specified amount or to minimize this cost;

6. To achieve a minimum level of total dollar contribution to cover joint fixed costs and earn profits for the firm or to maximize dollar contribution.

The rationale supporting each of these goals will be discussed. In addition, the mathematical representation for each goal will be given in the usual GP framework. Readers unfamiliar with the structure of GP constraints are encouraged to consult Lee [5].

Manpower level and productivity goals

On a departmental as well as a plant wide basis, the efficient utilization of manpower is an important consideration and criterion for evaluating performance. The following important decision areas will be optimized by the model so as to achieve several goals in management's hierarchy: hiring/firing decisions, rotating workers between departments, and the use of overtime. The following definitional constraint will be needed in several goals described below:

$$L_{k,t-1} + (NI_{kt} - ND_{kt}) = L_{kt}. \tag{1}$$

This recursive relationship shows the labor force size (in number of workers) in department k during period t will equal the labor force size of the previous period plus the net change in workers during period t.

Equation (2) quantifies management's goal for the number of productive hours available in each department k during each time period t:

$$T_k^1 (L_{k,t-1} - ND_{kt}) + T_k^2 NI_{kt} + T_k^3 O_{kt} + D_{1kt}^- - D_{1kt}^+ = \Sigma_i T_{ik}^4 P_{ikt}. \tag{2}$$

Management's goal for the production of each unit of product i in department k is given by the input parameter T_{ik}^4 which multiplied by the number of units to be produced (P_{ikt}—a decision variable in the model) and summed over all products, i, gives a target number of hours to be used in the department. The other

three input parameters T_k^1, T_k^2, and T_k^3 in this goal take learning curve effects into account by attaching a greater productivity coefficient for experienced workers (T_k^1) than for newly hired or transferred workers (T_k^2); similarly, possible differences in productivity of overtime hours (compared to regular time) are shown by multiplying the coefficient T_k^3 by the decision variable, O_{kt}, which gives the number of overtime hours to be worked in department k.

It should be noticed that the above goal assumes for simplicity that the learning curve is such that following one period of employment in a particular department, a new worker has attained the same level of proficiency as the more seasoned employees; of course, more complex assumptions could be incorporated into the model but the above illustrates a general two-stage learning effect. The deviational variables D_{1kt}^- and D_{1kt}^+ show respectively a deficiency and a surplus in available productive hours.

Job rotation and labor force stability goals

Considerable research cited in [3, Chapters 5 and 16] suggests that employee motivation, performance on the job, and satisfaction derived by workers are all enhanced when job enlargement or job rotation are implemented by the firm and when workers perceive a stable employment environment. The following two goals are incorporated into the model in order to provide at least a surrogate for job rotation and work-force stability.

The basic rationale for goal (3) is that management establishes a desired number of workers to be rotated among departments during period $t(M_t)$. Of course, discretion must be exercised in setting the level M_t so that workers and first-line supervisors do not view such rotation as disruptive rather than beneficial. Now, the number of transfers between departments will equal the *smaller* of either the total *net increases or* the total of the *net decreases* in the work force of all departments because it is assumed that all hiring and firing is done by a centralized personnel office within the plant, i.e., $\Delta_{1t} = \text{Min}\ (\Sigma_k\ \text{NI}_{kt}, \Sigma_k\ \text{ND}_{kt})$. Thus, workers not needed in one department will be transferred to another department where they are needed and will be released from the firm only if no department is in need of additional laborers. Similarly, new employees will be hired from the outside work force only if the number of workers transferred out of other departments is insufficient to meet the requirements of departments in need of additional workers. The goal is represented as:

$$\Delta_{1t} + D_{3t}^- - D_{3t}^+ = M_t \tag{3}$$

where D_{3t}^- and D_{3t}^+ are, respectively, underachieving or overachieving the established goal. The number of workers transferred between all pairs of departments is determined by the above goal, equation (1), the demand for various products, and subsequent goals which determine the optimal product mix and the optimal allocation of resources among departments so as to achieve the profitability goal as well as various cost control goals.

In addition to the above goal, workers also perceive greater job security if the firm does not make frequent and significant changes in the overall work force level. Further, the firm may feel that its image in the community and local labor force is enhanced through a conscious effort to maintain work force stability. Due to these considerations, which are relevant even apart from the economic costs involved in changes in aggregate work force level, the following goal is proposed:

$$\Delta_{2t} + D_{4t}^- - D_{4t}^+ = Q_t \tag{4}$$

where Q_t is the maximum desired fluctuation in aggregate work force level established by management; D_{4t}^- and D_{4t}^+ are, respectively, the number of workers less than or in excess of the desired maximum; and

$$\Delta_{2t} = |\Sigma_k NI_{kt} - \Sigma_k ND_{kt}|.$$

Due to centralized hiring and firing discussed above, the change in aggregate work force (Δ_{2t}) is the difference between the net increases and net decreases *among all departments*. As with goal (3), the present goal enters into the determination of the optimal values of the model's decision variables for the number of workers to be hired or fired in each period based on optimal product mix, departmental requirements, and cost considerations.

Cost minimization goals

Overall cost minimization is one of the most overriding criteria upon which aggregate production decisions are based. However, it is not

the sole criterion in the AOP problem and, more importantly, the various cost components of the overall cost function (i.e., production rate, work force level, and inventory) may be assigned significantly different importance by various firms. Thus, the new model has a distinct cost goal for each of the three major AOP cost components. This segregation provides flexibility to management in penalizing deviations from the various types of costs in that the cost goals can be placed on different priority levels which enables the consideration of a fuller range of tradeoffs among the cost components in model solution and through sensitivity analysis. If the firm desires to consider all costs simultaneously, they can be placed on the same priority level and, even here, they can be given different relative weights to show varying importance among the costs. The cost minimization goals shown below seek to find the optimal values of the decision variables by minimizing all relevant costs (both out-of-pocket and opportunity) according to a desired hierarchy and weighting scheme.

The following three cost minimization goals are established:

1. *Costs related to changes in work force size.* One set of relevant costs in AOP are those related to changing the size of the aggregate work force and inter-departmental transfers of workers. The following goal quantifies these costs:

$$C^1 \Delta_{3t} + C^2 \Delta_{4t} + C^3 \Delta_{1t} + D_{5t}^- - D_{5t}^+ = \mathrm{WFC}_t \quad (5)$$

where WFC_t is a target level of dollar cost incurred in period t in changing work force size and transferring workers between departments and D_{5t}^- and D_{5t}^+ are respectively a shortfall or excess of dollars spent relative to the target. The input parameters C^1, C^2, and C^3 are respectively the costs involved in hiring, firing and transferring an employee. It should be noted that Δ_{3t} shows the *net number of workers hired* and is found:

$$\Delta_{3t} = \begin{cases} \sum_k NI_{kt} - \sum_k ND_{kt} \text{ if positive} \\ 0 \text{ otherwise} \end{cases}$$

On the other hand, Δ_{4t} is the *net number of workers fired* in a given period and is equal to:

$$\Delta_{4t} = \begin{cases} \sum_k ND_{kt} - \sum_k NI_{kt} \text{ if positive} \\ 0 \text{ otherwise} \end{cases}$$

Thus, goal (5) is the major cost vehicle by which the decision variables related to changes in work force levels (Δ_{1t} thru Δ_{4t}) are optimized.

2. *Costs related to production rate.* Several relevant costs must be taken into account in determining the optimal production rate for the firm. These costs are incorporated in the following goal:

$$\sum_k \left\{ \sum_i (C_{ikt}^4 P_{ikt} + C_{ik}^5 H_{ikt} + C_{it}^6 (SC_{it} + X_{ikt}) + C_t^9 EI_{ikt} + C_{kt}^7 O_{kt} + C_{kt}^8 D_{1kt}^- \right\} + D_{6t}^- - D_{6t}^+ = PRC_t \quad (6)$$

where the necessary cost parameter inputs are the standard variable cost of producing one unit of product i in department k during period t (C_{ikt}^4), the cost on one unit of shrinkage of product i in department k (C_{ik}^5), the cost of subcontracting out one unit of product i (C_{it}^6), the cost per overtime hour in department k (C_{kt}^7) and the cost per idle-time hour in department k (C_{kt}^8). The value PRC_t is management's target level for production rate costs and the deviational variables are interpreted in the usual way. The variable H_{ikt} is an input which represents the shrinkage of work-in-process inventory of product i in department k. The remaining values in goal (6) are decision variables which are optimized by the model for each period; i.e., the number of units of product i to be produced in department k (P_{ikt}), the number of units of product i to be subcontracted out to increase finished good inventory (SC_{it}) or the work-in-process inventory in department k (X_{ikt}) and the number of overtime hours to be worked in department k (O_{kt}).

3. *Costs related to inventory level.* Inventory costs are another important component of total AOP costs and for finished goods include carrying costs, costs of shrinkage, and backorder costs. A target level for this cost is established for each time period t:

$$\Sigma_i (C_i^9 FG_{it} + C_i^{10} R_{it} + C_i^{11} BO_{it}) + D_{7t}^- - D_{7t}^+ = IC_t \quad (7)$$

where the cost parameter inputs required for this goal are the cost incurred for carrying one unit of product i (C_i^9), the shrinkage cost for one unit of product i (C_i^{10}), and the cost incurred for one unit of product i backordered per period (C_i^{11}); these costs should include

both out-of-pocket and opportunity costs. IC_t is management's target level for inventory cost and the values FG_{it} and BO_{it} are the number of units in finished good inventory of product i and the number of units of product i back-ordered respectively.

These values are computed from the inventory balance equation based on beginning inventory level, demand, production, and subcontracting. The amount of finished goods inventory shrinkage (R_{it}) can be a constant or a fixed percentage of the inventory level.

Of course, there are definite interrelationships among these three cost functions, and between each one and other goals formulated above. However, separating them provides greater flexibility to the firm in directing the focus of the optimization process.

Contribution margin goal

Besides the cost control goals, departments or overall facilities are often evaluated as profit centers wherein the total amount contributed to cover joint fixed costs and earn profits in each period is established as a target and used for performance evaluation. In addition, in the multi-product setting of this model, the dollar contribution per unit of each product will be used to arrive at the optimal product mix. The following goal incorporates this dimension:

$$\Sigma_i (SP_{it} - VC_{it}) S_{it} + D_{8t}^- - D_{8t}^+ = CM_t \qquad (8)$$

where SP_{it} is the selling price per unit of product i; VC_{it} is the standard variable cost per unit associated with the production and sale of product i in period t; S_{it} is the number of units of product i to be sold and shipped to customers in period t; CM_t is the dollar contribution goal with the deviational variables being interpreted in the usual manner. The values SP_{it}, VC_{it}, and CM_t are input parameters whereas S_{it} is a decision variable optimized by the model.

Economic constraints

Besides the managerial goals and policies described above using soft constraints, there are several relevant economic constraints which define relationships and resource limitations that the firm faces in its AOP process. These latter constraints determine the feasible alternatives available to the firm as it strives to achieve its multiple goals discussed above.

The major economic constraints incorporated in the model are:

1. A budget constraint quantifying a strict limitation on cash outflows on a departmental or plant-wide basis;

2. A constraint which shows other limited productive resources required to produce various products;

3. An inventory balance equation for work-in-process units;

4. An inventory balance equation for all finished products; and

5. Constraints showing upper and/or lower limits on any decision variable in the model.

Because these constraints are somewhat common in usage, they will not be specifically formulated here. (See [5].)

IMPLEMENTATION AND CONCLUSIONS

During the implementation stage, all GP models become unique to the subject firm. Each company has its own hierarchy of objectives and preferences concerning tradeoffs among goals. These different preferences and objectives are reflected in the magnitude of the managerial goals established, the priority level selection as well as the relative weights assigned to deviational variables within those priority levels.

The first question to be resolved in order to implement the proposed GP model is the scope or focus which the model is to take. For example, at the facility level, AOP is usually related to a single economic sector; whereas, at the corporate level, fluctuations in demand in a single economic sector are insulated from exposure as a result of the corporation's product mix.

The time element of this GP AOP model is of considerable importance during the initial phases of implementation. Model administrators should consider learning curve effects, production lead-time, shrinkage rates, and subcontract procurement lead-time as the primary limiting factors in selection of the proper time

period duration. Furthermore, a planning horizon based on factors limited primarily by capital equipment acquisition lead time (the amount of time required to increase facilities capacity) must be selected. However, more horizons tend to be annual corresponding to budgeting and related financial decisions. Another consideration of the model is time references as limited from the short duration side. Since there is a period of time over which all factors of production are fixed, it would be meaningless to design a model with reference less than that period.

Frequently multiproduct manufacturing operations utilize intermediate products as common input for the production of many finished goods. During the implementation stage of the model, it will become necessary to identify those stages of manufacture where products become unique (i.e., the split-off points). Continuous process manufacturing concerns in particular must devote considerable effort to this stage. On the other hand, job shops tend typically to manufacture goods which require different sequences of operations. Identification of this sequence for each product is mandatory as a result of the model's accounting for interdepartmental transfers.

Certain conversion constants are required in the model. These constants should reflect direct (out-of-pocket), indirect (ramification), and opportunity costs. Indeed great care must be exercised in the accurate evaluation of these factors.

Each management team has its own set of corporate or plant objectives in terms of long-run plans, productive capacities, product lines, etc. Further, each manufacturing system operates subject to a cost structure which results from those plans, capacities, product lines and so forth. As a consequence the model needs to be specifically tailored to reflect these aspects. Certainly the time chain effects of those constraints are obvious to the professional manager. But, the priority levels for the various goals, the quantification of certain among those goals and the interrelationships among those goals warrant the most serious consideration affordable. It is in this area of implementation that the most time must be spent in order for the final model to accurately replicate the reality of the situation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the helpful comments of Dr John F Muth, Indiana University; Dr Gene K Groff, Georgia State University; Dr Michael F Pohlen, University of Delaware; Dr William Berry, Indiana University; Dr Elwood S Buffa, UCLA; and Dr Laurence J Moore, VPI. Of course, any remaining errors of omission or commission are the exclusive responsibility of the authors.

REFERENCES

1. Buffa ES & Taubert WH (1972) *Production—Inventory Systems: Planning and Control* (Rev. Edn.) Richard D Irwin, Homewood, Illinois.
2. Goodman DA (1974) Goal programming approach to aggregate planning of production and work force. *Mgmt Sci.* **20**, 1569–1575.
3. Groff GK & Muth JF (1972) *Operations Management: Analysis for Decisions.* Richard D Irwin, Homewood, Illinois.
4. Jaakelainen V (1969) A goal programming model of aggregate production planning. *Swedish J. of Econ.* **2**, 14–27.
5. Lee SM (1972) *Goal Programming for Decision Analysis.* Auerbach, Philadelphia, Pennsylvania.

ADDRESS FOR CORRESPONDENCE: Thomas J Hindelang, Esq, Department of Finance and Statistics, College of Business and Administration, Drexel University, Philadelphia, Pennsylvania 19104, USA.

MEMORANDA

Vol. 6
1978

BLANK
PAGE

Vol.
1978

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 275–276
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0275$02.00/0

An Equivalent Value Transform

THIS paper develops the concept of a pseudoconcave transform which is useful in dealing with decision analysis problems where the objective function is specified in terms of equivalent value or certainty equivalent instead of expected utility. Using equivalent value may allow additional economic insights that can be obscured by the expected utility.

PSEUDOCONCAVE TRANSFORM

The concept of pseudoconcavity is discussed by Mangasarian [5, p. 141]. Let $g(\cdot)$ be a numerical function defined on some open set in $\mathbf{R}^n$ containing the set Γ. We say $g(\cdot)$ is pseudoconcave at $\bar{\mathbf{x}} \in \Gamma$ (with respect to Γ) if it is differentiable at $\bar{\mathbf{x}}$ and that

$$\nabla g(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \leq 0 \Rightarrow g(\mathbf{x}) \leq g(\bar{\mathbf{x}}), \forall \mathbf{x} \in \mu \qquad (1)$$

$g(\cdot)$ is said to be pseudoconcave on Γ if it is pseudoconcave at each $\mathbf{x} \in \Gamma$. We will assume that $g(\cdot)$ has first and second partial derivatives. We will also say that $g(\cdot)$ has a pseudoconcave transform on Γ if there exists a strictly increasing function $w(\cdot)$ such that $w[g(\mathbf{x})]$ is pseudoconcave on Γ, we will assume that $w(\cdot)$ has first and second derivatives.

At this point we will show that $g(\mathbf{x})$ is pseudoconcave if and only if $w[g(\mathbf{x})]$ is pseudoconcave. We begin the proof by noting that if $w[g(\mathbf{x})]$ is pseudoconcave then, for any $\bar{\mathbf{x}} \in \Gamma$, we know.

$$w'[g(\bar{\mathbf{x}})]\nabla g(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \leq 0 \Rightarrow w[g(\mathbf{x})] \leq w[g(\bar{\mathbf{x}})]. \qquad (2)$$

Since $w'[g(\bar{\mathbf{x}})] > 0$, it obviously follows that $g(\cdot)$ is pseudoconcave at $\bar{\mathbf{x}}$ and thus $g(\cdot)$ pseudoconcave on Γ. To complete the proof we need to show that if $g(\mathbf{x})$ is pseudoconcave then $w[g(\mathbf{x})]$ is pseudoconcave. If $g(\mathbf{x})$ is pseudoconcave then, by the obvious application of $w'[g(\bar{\mathbf{x}})] > 0$ on (1), (2) immediately follows. This completes the proof.

THE EQUIVALENT VALUE FORMULATION

Consider a decision maker (DM) who makes choices by maximizing his expected utility in accordance with the von Neumann–Morgenstern utility hypothesis [6]. The DM is uncertain about the relationship between his feasible action choices, $\mathbf{X}$, and the relevant outcome in the future as given in terms of a money amount $g(\mathbf{x})$. Using z to index this uncertainty where $\mathbf{Z} = \{z\}$, or let $g(\mathbf{x};z)$ represent the money amount for action $\mathbf{x}$ and state z. We will assume $g(\mathbf{x};z)$ is concave on $\mathbf{X}$ for all $z \in \mathbf{Z}$. This assumption implies a non-increasing rate of return.

Since the money amount, $g(\mathbf{x};z)$, is uncertain, the DM will choose $\mathbf{x}$ so as to maximize expected utility. So, letting $u(\cdot)$ denote the utility function, the DM has the following problem

$$\text{Max}_{\mathbf{x} \in \mathbf{X}} U(\mathbf{x}) \text{ where } U(\mathbf{x}) = E_z u[g(\mathbf{x};z)]. \qquad (3)$$

We assume that the DM is not a risk preferer and thus $u(\cdot)$ is concave. Moreover we assume that $u(\cdot)$ is strictly increasing on money.

An equivalent problem would replace $U(\mathbf{x})$ with $\phi(\mathbf{x})$ in (5) where $\phi(\mathbf{x})$ is a dollar amount such that $u[\phi(\mathbf{x})] = U(\mathbf{x})$. Although the two formulations are equivalent, in the sense of having the same solutions, the equivalent value is preferred in many cases, as indicated by Berhold [2], because it provides a better conceptual reference than the abstract notion of expected utility.

While equivalent value and expected utility formulations both have the same solution, we can show that, while $U[\phi(\mathbf{x})]$ is concave on $\mathbf{X}$, $\phi(\mathbf{x})$ is not necessarily concave on $\mathbf{X}$. This result is counterintuitive but it is quite easy to illustrate through a counter example.

Let

$$u(g) = \begin{Bmatrix} 10g \\ 2g + 80 \end{Bmatrix} \quad \text{for} \quad g \begin{Bmatrix} < \\ \geq \end{Bmatrix} 10$$

with the following payoff functions

$$g(x;1) = x$$
$$g(x;2) = x + 6$$

where $p(z = 1) = 1/2$ and $p(z = 2) = 1/2$. We then get expected utility:

$$U(x) = E_z u[g(x;z)] = \begin{Bmatrix} 10x + 30 \\ 6x + 46 \\ 2x + 86 \end{Bmatrix} \text{when} \begin{Bmatrix} x < 4 \\ 4 \leq x < 10 \\ x \geq 10 \end{Bmatrix}$$

from which we get $\phi(x)$; i.e.

$$\phi(x) = \begin{Bmatrix} x + 3 \\ 0.6x + 4.6 \\ 3x - 17 \\ x + 3 \end{Bmatrix} \text{for} \begin{Bmatrix} x < 4 \\ 4 \leq x < 9 \\ 9 \leq x < 10 \\ x \geq 10 \end{Bmatrix}.$$

It is quite clear that $\phi(x)$ is not concave even though $U(x)$ is concave.

We can make two observations based on the results in the first section. First, the direct observation that $U(\mathbf{x})$ is pseudoconcave if and only if $\phi(\mathbf{x})$ is pseudoconcave. Second, as a minor extension, if $U(\mathbf{x})$ is concave then it is pseudoconcave [5, p. 147] and thus $\phi(\mathbf{x})$ is also pseudoconcave as before.

DISCUSSION

In assuming that the decision maker is not a risk preferer, we are not necessarily being very restrictive. Decreasing returns to scale is a common assumption in economic contexts [1, pp. 261–262] and this generally implies risk aversion. In the model of contractural sharing rules [3] we also see that risk preference leads to unbounded solutions which also suggests that the decision makers are not risk preferers. Thus risk preference can be viewed as an anomaly.

With the concept of a pseudoconcave transform we are able to use the equivalent value formulation without complicating the analysis. This is particularly useful in problems of constrained optimization. To illustrate, if a utility maximization problem uses equivalent value

to formulate a nonlinear programming problem then the marginal tradeoffs or shadow prices can be expressed in terms of marginal equivalent values as opposed to marginal expected utilities. Thus, without complicating the analysis, shadow prices would be given a more obvious and less abstract economic interpretation. A discussion of the economic interpretations of the nonlinear programming formulation is provided by [4, pp. 186–203].

Aside from the specific benefits that have been discussed, we might note that alternative formulations are of general use because they offer alternative ways of looking at the same problem. Thus, when alternative formulations are available it becomes routine to look at each formulation to aid in interpretation or to aid in the solution. Perhaps the best example of this is the primal and dual problem in mathematical programming and game theory. So one might suspect that additional benefits of the pseudoconcave transform might develop as this concept is applied in other contexts.

SUMMARY AND CONCLUSIONS

The existence of pseudoconcavity is very useful because the necessary conditions for an optimal solution will also be sufficient conditions. This useful property is extended in this paper by introducing the concept of a pseudoconcave transform.

This concept is particularly useful in the equivalent value formulation of the decision analysis model because pseudoconcavity on expected utility provides pseudoconcavity on equivalent value. The equivalent value formulation is preferred because it can provide the basis for insights into the problem that might be obscured by the expected utility formulation. This is particularly true in a nonlinear programming formulation where expected utility might obscure some of the economic interpretations that are available under the equivalent value formulation.

REFERENCES

1. Allen RGD (1964) *Mathematical Analysis for Economists.* St. Martin's, New York.
2. Berhold M (1971) Applications of a risk aversion concept. *Dec. Sci.* **2**(2), 129–140.
3. Berhold M (1976) Contractual sharing rules. *Omega* **4**(4), 487–488.
4. Dorfman R, Samuelson PA & Solow RM (1958) *Linear Programming and Economic Analysis.* McGraw-Hill, New York.
5. Mangasarian OL (1969) *Nonlinear Programming.* McGraw-Hill, New York.
6. Von Neumann J & Morgenstern O (1947) *Theory of Games and Economic Behavior*, 2nd edn. Princeton University Press, New Haven.

Prof. MH Berhold

(*November* 1977)

Georgia State University
Department of Quantitative Methods
University Plaza
Atlanta
Georgia 30303
USA

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci.. Vol. 6. No 3. pp. 277–279
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0277$02.00/0

A Proposal for Classification of Inventory Systems

INTRODUCTION

THERE is an abundance of literature on inventory problems but as yet no generally acceptable system of classification of inventory systems. In contrast, there is a generally accepted standard classification of queueing systems based on the proposal of Kendall [4], modified by Lee [5]. Attempts by various authors to classify inventory systems have so far considered only operating aspects of policies relevant to the systems; Eilon and Lamptou [2] Newberry [7], Naddor [6], Fabrycky and Banks [2] and Aggarwal [1] have proposed such classifications. Extending Naddor's definition to include procurement cost, an "inventory system" is a system in which carrying, shortage, replenishment and procurement costs are significant and will be denoted as type 1, 2, 3 and 4 costs respectively, hence C_1, C_2, C_3, and C_4. An "inventory problem" is one of making optimal decisions with respect to time, quantity, source and mode of procurement of inventory and frequency of inventory review, whereas an "inventory policy" is an operating policy resulting from a modelling and solution process.

In this paper a system of classification is proposed which includes all major aspects of inventory systems. In this proposal, previous classifications have been synthesized to obtain a more comprehensive scheme.

PROPOSED CLASSIFICATION OF INVENTORY SYSTEMS

This proposal tries to encompass every aspect of an inventory system by describing it in a sequence of four aspects, namely, structure of the system, environmental parameters, operating policies and inventory-related costs. It is hoped that this scheme will cover all relevant information pertaining to the system in compact form. Thus, the basic format proposed is:

Structure/Environmental parameters/Operating Policies/Costs

For each component of the total classification scheme, the following notations are used:

1. *Structure*

This refers to the basic elements of the system including the number of items, number of sources of supply, number of echelons and locations at which an item is stocked before reaching actual demand points.

The proposed sequence to define the structure is:

Number of items, number of sources of supply, number of echelons and their configuration.

Notation is as follows:

SI — Single Item
MI — Multi-Item
SS — Single Source
MS — Multi-Source
SE — Single Echelon
ME — Multi-Echelon with subscripts for series: *p* for parallel and *a* for hybrid or arborescence structure.

Thus a typical inventory system structure with single item, multi-source and parallel configuration of multi-echelon nature will be represented in the proposed scheme as (SI, MS, ME_p).

2. *Environmental parameters*

Four major parameters in a prescribed sequence are proposed to describe the environmental constraints within which an inventory system has to function. These are the nature of demand variation, replenishment time distribution, shelf life and size of back-logging permitted. If no back-logging is allowed then the case is one of lost sales; if back-logging is finite (R), then there is an upper bound to back-logging beyond which lost sales conditions operate and if back-logging is infinite (∞) then there is no restriction but lost sales may still occur.

Notation is as follows:

D — Deterministic demand, replenishment time or shelf life.
M — Poison demand, exponential replenishment time or shelf life.
E_1 — Erlang demand with shape parameter 1; shape parameters for Erlang replenishment times and shelf life may be denoted by k and j respectively. Instantaneous replenishment (infinite rate) and non-perishable nature of items can be denoted by ∞.
G_1 — General independent demand distribution.
G — General replenishment time or shelf life distribution.
W — Weibull shelf life distribution.
CP — Compound Poisson demand indicating an erratic demand pattern with occasional high demands interspersed with low demands.
R — Size of back-logging permitted.

Thus a typical environment for an inventory system with Erlang demand, exponential replenishment time and general shelf life distribution with complete back-logging can be symbolised by (E_1, M, G, ∞). Similarly, (D/D, M/D, ∞, 0) will describe a multi-echelon system with deterministic demand at all echelons, exponential replenishment time at the top echelon and deterministic replenishment time at lower echelons for a non-perishable item without back-logging.

3. *Inventory Policies*

Major inventory decisions are order size (q), reorder

point (s), review period (t) and order level (S). If any of these are prescribed, then this will denoted by a subscript p for instance, q_p indicates a prescribed order size of q and thus order size is not a decision variable. If these parameters are non-stationary over time, then this will be denoted by a subscript i. For example, q_i may indicate a variable order size for increasing demand or inflationary conditions. For multi-echelon systems, inventory policies for various echelons may be denoted using these symbols in a similar arrangement to that described for environmental parameters. Using Naddor's [6] notations, as far as possible, the following major inventory policies are identified:—

1. (t, S) place order every review period to bring inventory up to order level S.
2. (s, q) place order of size q when inventory falls to the reorder point s.
3. (s, S) place order to bring inventory up to order level S when stock falls to the reorder point s at review period t.
4. (s, $\propto$ q)(s, q) policy with an average fraction of order size realized (for $\propto >$ or $\leqslant 1$).
5. (s, C, S) co-ordinated or joint replenishment policy.
6. (t_p, S) prescribed scheduling period order level policy.
7. (s_p, q) prescribed reorder point, order size policy.
8. (s, q_p) prescribed order size, reorder point policy.
9. (s-1, s) place order of size one when each demand of size one occurs.
10. (s_j, q) multiple-reorder point, order size policy.
11. (t_i, s_i)(t, s_i) or (s_i, q_i) policies for increasing demands or inflationary trends. (Subscript i denotes variation in inventory decisions as a function of time).
12. (s, S)/(s-1$_j$, s_j) = multi-echelon system with (s, S) policy at the top echelon and (s-1$_j$, s_j) for the lower echelons.
13. (s, q)/a wave-replenishment policy in a multi-echelon system with (s, q) policy at the top echelon and immediate distribution of inventory to all lower echelons with a hold-back at the top echelon. Hold-back factor $a < 1.00$.

There are a number of other inventory polices in use and an exhaustive list of these policies is difficult to prepare. However, it is felt that the above notations include all the major inventory policies employed in practice.

4. *Inventory-Related Costs*

It is proposed to denote the relevant costs in the sequence they are defined, omitting those costs which are irrelevant in decision making. Thus (C_1, C_3) defines a system in which only carrying and replenishment costs are subject to control. Procurement cost C_4 will be relevant in conditions where price breaks, inflation, transportation costs and source-dependent prices are significant factors and have a bearing on decisions regarding the inventory system; in other situations it may be dropped. A description (C_1, C_2, C_3, C_4) indicates that all four costs are subject to control.

Other important aspects of cost optimization are the planning period and whether or not costs are discounted. A description (C_1, C_2, C_3, C_4) is proposed to denote a single period optimization problem and if N-period optimization of costs is considered then this will be indicated by $N(C_1, C_2, C_3, C_4)$. Note that N could be $\propto$ to indicate an infinite planning horizon. For discounted costs with discount rate i, the costs could be denoted N^i (C_1, C_2, C_3, C_4). Thus $\propto 1$ (C_1, C_2) describes an inventory system with infinite planning horizon for discounted carrying and shortage costs. In some cases, particulary with the (s, S) policy, there may not be an ordering cost but there is a review cost; both are included in the replenishment cost C_2.

SOME TYPICAL EXAMPLES

1. (SI, MS, SE)/(D, D, $\propto$, R)/(s, q)/$N^i(C_1, C_2, C_3, C_4)$

This represents a system with single item, multi-source, single echelon structure with deterministic demand and replenishment times of a non-perishable item with finite size of back-logging employing an order size, reorder point policy with an objective of minimising N-period discounted costs.

2. (MI, SS, SE)/(GI, G, M, $\propto$)/(s, S)/$\propto(C_1, C_2, C_3)$

This represents a multi-item, single source single echelon structure with general demand and replenishment times and exponential shelf life distribution with complete back-logging employing an (s, S) policy with an objective of minimising the first three inventory-related costs over an infinite planning horizon.

3. (SI, SS, ME_p)/(GI/M, D/D $\propto$, $\propto$)/(s, S)/(s-1, s)/(C_1, C_2, C_3, C_4)

This represents a single item, single source system with parallel configuration of multi-echelon nature with general demand at the top echelon and Poisson demand at the lower echelons and deterministic replenishments with complete back-logging of a non-perishable item using an (s, S) policy at the top echelon and (s-1, s) policy at the lower echelons to minimize single period costs of all types.

CONCLUSIONS

The method proposed for classification of inventory systems has the advantage that it includes far more information than previously suggested schemes. This attempt integrates the earlier efforts and borrows heavily from the work of Kendall, Naddor, Fabrycky and Banks. Use is made of the notations of these authors to facilitate quicker understanding of the scheme proposed in this paper. If it is adopted, the proposed scheme could become a relatively easy and comprehensive way of classifying inventory systems.

It is worthwhile mentioning that at a first glance the proposed scheme appears to be complex, but it is not really so if the basis of classification is first understood. Since the proposed scheme conveys all major information about characteristics of inventory systems it is bound to be more difficult than Kendall's notation for queuing systems. It must be admitted that the proposal does not include all the parameters in an exhaustive way as this is not feasible for any classification scheme; but it does include what the authors consider to be the major information about the inventory system. Finally, it must be emphasized that the scheme should be viewed as a feasible proposal and may leave scope for further refinement and improvement so that a generally acceptable method of classification may eventually be evolved. The authors would welcome any such suggestions for modification of the scheme proposed.

REFERENCES

1. AGGARWAL SC (1973) A review of current inventory theory and applications. *Int. J. Prod. Res.* **12**(4), 443–482.

Vol. 6
1978

2. EILON S & LAMPTON W (1968) Inventory Control Abstracts 1953–1965. Oliver and Boyd, Edinburgh.
3. FABRYCKY WJ & BANKS J (1970) *Procurement and Inventory Systems*. Reinhold, New York.
4. KENDALL DG (1951) Some problems in the theory of queues. *J. R. Statist. Soc.* **B-13**(2), 151–185.
5. LEE AM (1966) *Applied Queuing Theory*. Macmillan, London.
6. NADDOR E (1966) *Inventory Systems*. John Wiley, New York.
7. NEWBERRY TL (1960) A classification of inventory control theory. J. Ind. Engng XI, 391–397.

Prof. RH Hollier
Dr Prem Vrat

The University of Manchester Institute of Science and Technology
Department of Management Sciences
PO Box 88
Manchester M60 1QD
UK

Indian Institute of Technology
New Dehli-110029
India

(*December* 1977)

ol. 6
978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 280–281
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0280$02.00/0

Conservatism in Bayesian Information Processing

THE PURPOSE of this study was to investigate how individuals revise probabilities in two situations, the bookbag and pokerchip paradigm and a consumer task. In both situations, subjects were asked to assess likelihood probabilities as well as prior and posterior probabilities so that estimated posteriors could be compared with the posteriors computed using Bayes' theorem and the subjects' prior and likelihood probabilities.

METHODOLOGY

Sixty-one paid undergraduate students participated in the tasks. In the first task, subjects were shown two boxes of pokerchips and were told one contained 70 red chips and 30 blue chips and the other 70 blue and 30 red; one box was randomly selected by flipping a coin. After assessing prior probabilities that the predominantly red or predominantly blue box was chosen, subjects were shown a sample of seven chips (four red and three blue) taken from the selected box and subjects were asked to assess the conditional probabilities of the sample information.

In the second task, subjects were informed that only two stores in the town carried a particular brand of shoe and were given information about the shoes and the stores; then they were asked to assess the probability that each store had a wider selection (larger number of styles). Thus,

H_A = the event that Store A has the wider selection,

and

H_B = the event that store B has the wider selection.

Next, subjects were shown a hypothetical advertisement announcing that Store A had just received a shipment of shoes and were asked to assess the likelihood, $P(D|H_A)$, that they would see such information if Store A had the wider selection. Then they were asked the same question given that Store B had the wider selection. Finally, they were asked again to assess the probability that each store had the wider selection, $P(H_A|D)$ and $P(H_B|D)$.

HYPOTHESES

Three hypotheses were developed: (1) subjects' posterior estimates in both tasks are consistent with posteriors calculated using Bayes' theorem; (2) subjects show no general bias or, since the bias has usually been in one direction, subjects are not conservative in their revisions; and (3) subjects perform better (the posterior probabilities calculated using Bayes' theorem will be closer to their posterior estimates) in the consumer task than in the traditional task.

FINDINGS

The dependent variable for proximity of estimated posterior to predicted posterior is the difference score between the subjective log likelihood ratio and the Bayesian log likelihood ratio (SLLR − BLLR), where

$$\text{SLLR} = \text{Log}(\overset{*}{P}(H_1|D)/\overset{*}{P}(H_2|D)) - \text{Log}(P(H_1)/P(H_2)),$$

and

$$\text{BLLR} = \text{Log}(P(H_1|D)/P(H_2|D)) - \text{Log}(P(H_1)/P(H_2)).$$

The subjects estimated posteriors are denoted by $\overset{*}{P}(H_i|D)$ and the posteriors computed via Bayes' theorem are $P(H_j|D)$.

Table 1 shows distributions of difference scores in various situations. None of the distributions is tightly bunched around a score of 0.0, which would have occurred if subjects' posteriors were correctly predicted by the model's estimates. In the bookbag task, subjects were generally conservative and tended to be more conservative when theoretical rather than subjective likelihoods were used; however, subjects did not exhibit conservatism in the consumer task. When the three distributions were compared using the Kolmogorov–Smirnov test for goodness of fit, they were found to be significantly different beyond the 0.01 level.

Our second hypothesis was that subjects would show no general bias; they would not underweight the impact of information in their posterior estimates as com-

TABLE 1. DISTRIBUTIONS OF DIFFERENCES SCORES

	Range of scores								
Iteration	< -1.0	< -0.6 ≥ -1.0	< -0.3 ≥ -0.6	< -0.001 ≥ -0.3	<0.001 ≥ 0.001	<0.3 >0.001	<0.6 ≥ 0.3	<1.0 ≥ 0.6	≥ 1.0
Bookbag, subjective likelihoods	3* 0.048	8 0.131	11 0.180	20 0.328	5 0.082	6 0.099	4 0.066	4 0.066	0 0.000
Bookbag, theoretical likelihoods	4 0.066	15 0.246	25 0.426	5 0.082	9 0.148	0 0.000	2 0.033	1 0.016	0 0.000
Consumer Task	0 0.000	3 0.048	7 0.115	6 0.099	3 0.049	14 0.230	9 0.148	11 0.180	8 0.131

* The top number is the absolute frequency; the bottom number is the proportion of the 61 subjects in that range.

TABLE 2. T-VALUES FOR SECOND HYPOTHESIS

Iteration	Mean difference score	Standard deviation	t Value	Degrees of freedom
Bookbag, subjective likelihoods	−0.225	0.519	−3.38*	60
Bookbag, theoretical likelihoods	−0.421	0.378	−8.70*	60
Consumer task	0.377	0.776	3.79*	60

* $P < 0.001$.

TABLE 3. T-VALUE FOR THIRD HYPOTHESIS

Iterations compared	Mean difference score	Standard deviation	t Value	Degrees of freedom
Bookbag (subjective) and Consumer task	−0.1798	0.7397	−1.898*	60

* $0.025 < P < 0.05$.

pared to their assessments of the likelihoods (conservatism), nor would they overweight the impact of information in their posterior estimates. The findings in Table 2 indicate that subjects tend to be conservative in the traditional task, but not in the consumer task; results from bookbag and pokerchip paradigms cannot be generalized to other decision situations.

For the third hypothesis, we expected subjects to be more consistent in the consumer task than in the pokerchip task. "More consistent" was operationally defined as having a difference score closer to 0; absolute deviations of subjects' scores from 0.0 were compared using paired observations. Thus:

$$H_0: \mu_d \geq 0$$

versus

$$H_A: \mu_d < 0$$

and the statistic of interest is

$$\bar{d} = \sum_{i=1}^{n} (|D_{1i}| - |D_{2i}|)/n$$

where D_{1i} refers to the bookbag and pokerchip task and D_{2i} refers to the consumer task.

From findings in Table 3, subjects performed significantly better in the bookbag and pokerchip task than in the consumer task; the opposite result would occur if (1) subjects use Bayes' theorem as their combinatorial rule and (2) the selection task is more familiar to the subjects, enabling them to adapt their combinatorial rule to the setting more easily. One or both of these assumptions was invalid.

CONCLUSIONS

In general, our subjects were conservative in the traditional scenario; they did not revise their prior probabilities as much as was prescribed by Bayes' theorem. However, conservatism was not a generalized finding in the consumer-oriented task, so conservatism appears to be a function of the tasks rather than a function of the subjects themselves.

It had been predicted from previous research ([1,2]) that assessments in the consumer task would be more consistent with predictions from Bayes' theorem than those made in the bookbag and pokerchip task. The results of this study do not support that contention, since subjects were inconsistent in both situations but performance in the consumer task was worse than in the bookbag task.

No empirical support was found for a conclusion that added familiarity with the task would result in better consistency between subjects' assessments and those predicted by Bayes' theorem; the concept of conditional probability seems difficult for subjects to grasp in a real-world situation.

REFERENCES

1. DUCHARME WM & PETERSON CR (1968) Intuitive inference about normally distributed populations. *J. Expl. Psychol.* **78**, 269–275.
2. WINKLER RL & MURPHY AH (1973) Experiments in the laboratory and the real world. *Orgl Behav. Hum. Perf.* **10**, 252–270.

Prof. TV O'Brien
JW Gentry

(*November* 1977)

Kansas State University
College of Business Administration
Calvin Hall
Manhattan
Kansas 66506
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 282–283
Pergamon Press Ltd 1978. Printed in Great Britain

0305-0483/78/0701-0282$02.00/0

An Approach to Linear Modeling with Dispersion Surrogates

A COMMON representation of a linear programming problem is:

$$\begin{aligned} &\max \mathbf{c}\cdot\mathbf{x} \\ &\quad \mathbf{A}\mathbf{x} \le \mathbf{b} \quad \mathbf{A} \text{ is } mxn \\ &\quad \mathbf{x} \ge \mathbf{0} \end{aligned} \qquad \text{(LP)}$$

It is well-known that if an optimal solution to (LP) exists then there exists an extreme point optimal solution with at most m variables different from zero. If m is very much smaller than n, then a large number of the variables will be zero. Such a result may produce certain misgivings for a diversified firm which manufactures many different products or the government agency empowered to allocate funds to a large number of political districts. Considerations for long range growth, stability or even political pressures may indicate a requirement for more homogeneity of the decision variables.

These observations could be interpreted as simply sufficient reasons why a linear model is not appropriate, i.e. the decision maker may be unhappy with an optimal solution with just a few variables different from zero. One might argue, though, that the linear model would remain reasonable in these circumstances if a positive bound was prescribed for each of the variables. Alternatively, one might include explicit nonlinear relationships in the model that penalize variations in the relative size of the decision variables. For example, in the classic aggregate planning model, where x_t is the production scheduled for period t, one often includes a production rate change cost of the form $\mathbf{c}\cdot(\mathbf{x}_t - \mathbf{x}_{t+1})^2$.

In the absence of any specific guidelines for these lower bounds or in view of the multitude of different values or costs that might be assigned we will present an alternative methodology for adapting the linear programming problem. The motivation for this methodology hinges on the observation that in general any feasible extreme point to (LP) is contained in several of the bounding hyperplanes $x_j \ge 0$, i.e. the non-negative orthant define by R^n_+. The discussion seems to indicate the need for a rationale of choosing a decision vector that lies somewhat "deeper" into the interior of R^n_+.

One vector that is deep into the interior of the cone R^n_+ is the vector $\mathbf{e}$, which contains one as each component. A radical concession to the goal of equalizing the size of all the decision variables would be the formulation of (LP-e) as follows:

$$\begin{aligned} &\max \mathbf{c}\cdot\mathbf{x} \\ &\quad \mathbf{A}\mathbf{x} \le \mathbf{b} \\ &\quad \mathbf{x} - \lambda\mathbf{e} = \mathbf{0} \\ &\quad \mathbf{x}, \lambda \ge \mathbf{0} \end{aligned} \qquad \text{(LP-e)}$$

In (LP-e) the decision vector is restricted in that all components must be the same.

Actually one could consider any vector $\bar{\mathbf{x}}$ in the interior of R^n_+ and not just the vector $\mathbf{e}$. (LP-$\bar{\mathbf{x}}$) is formulated as:

$$\begin{aligned} &\max \mathbf{c}\cdot\mathbf{x} \\ &\quad \mathbf{A}\mathbf{x} \le \mathbf{b} \\ &\quad \mathbf{x}, \lambda \ge \mathbf{0} \\ &\quad \mathbf{x} - \lambda\bar{\mathbf{x}} = \mathbf{0} \end{aligned} \qquad \text{(LP-}\bar{\mathbf{x}}\text{)}$$

and one could consider all intermediaries between (LP) and (LP-$\bar{\mathbf{x}}$). Such a procedure would be realized, for example, if the variables are restrained to be approximately in a fixed proportion or if a given solution $\bar{\mathbf{x}} \ge \mathbf{0}$ is in existence prior to analysis via linear programming optimization. Note that the problem (LP-$\bar{\mathbf{x}}$) can be normalized by scaling each variable, $x'_i = x_i/\bar{x}_i$ and substituting $x_i = \bar{x}_i \cdot x'_i$ into (LP-$\bar{\mathbf{x}}$). In this manner (LP-$\bar{\mathbf{x}}$) can always be converted into normalized form (LP-e).

Consideration of the problem (LP-$\bar{\mathbf{x}}$) could be for several reasons. Suppose, because of changing prices and demands, that a firm is forced to modify its product mix that currently is the vector $\bar{\mathbf{x}}$. Rather than generating a radically different product mix, $\bar{\mathbf{y}}$, by resolving (LP) it may be more advisable to restrict the new vector $\bar{\mathbf{y}}$ to be in the "neighborhood" of the old vector $\bar{\mathbf{x}}$. One may want, for a particular time frame, to control the rate of change from $\bar{\mathbf{x}}$ to $\bar{\mathbf{y}}$, i.e. in early periods allow only a small deviation from $\bar{\mathbf{x}}$. In general, of course, (LP-e) would not possess a feasible solution, and something intermediate between (LP) and (LP-e) should be in order. The intermediary family of problems between (LP) and (LP-e) analyzed in this note is defined as:

$$\begin{aligned} &\max \mathbf{c}\cdot\mathbf{x} \\ &\quad \mathbf{A}\mathbf{x} \le \mathbf{b} \\ &\quad \mathbf{x} \in K_\alpha \end{aligned} \qquad (P_\alpha)$$

Where

$$K_\alpha \equiv \left\{ \mathbf{x} \in R^n_+ \,|\, x_j \sum_{i=1}^{n} x_i \ge \alpha,\ j = 1, 2, \ldots n \right\}.$$

In (P_α), the variable x_j accounts for at least α percent of the total activity Σx_i. Observe that when $\alpha = 0$, (P_α) becomes (LP) and, when $\alpha = 1/n$, (P_α) becomes (LP − e).

Now let $p(\alpha)$ be the optimal value of (p_α). The function $p(\alpha)$ is non-decreasing on its respective domain $[0, 1/n]$, since $\alpha_1 \le \alpha_2$ implies $K_{\alpha_2} \subseteq K_{\alpha_1}$. Also, assume that (p_α) is feasible at $\alpha = \bar{\alpha}$.

If $\mathbf{A}\mathbf{x} \le \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$ is bounded, $\bar{\mathbf{x}}$ is a solution to this linear system and $\bar{\mathbf{x}}$ is in the interior of $K_{\bar{\alpha}}$, then $p(\alpha)$ is continuous at $\bar{\alpha}$.

In certain cases one may be interested in α^*, the maximum feasible value of α, along with a solution that corresponds to α^*. Rather than obtain α^* as a by-product of a parametric analysis of (P_α), which may involve a very large number of basis changes and hence be extremely time consuming, one may determine α^* directly by solving a certain linear programming problem. The nominal problem to be solved is the nonlinear

Vol. 6
1978

program:

$$\begin{aligned} &\sup\ \alpha \\ &\mathbf{Ax} \le \mathbf{b} \qquad \text{(NLP)} \\ &x_j \Big/ \sum_{i=1}^{n} x_i - \alpha \ge 0,\ j = 1, 2, \ldots n \\ &x_j \ge 0 \end{aligned}$$

(We assume that $\mathbf{x} = \mathbf{0}$ is not feasible for $\mathbf{Ax} \le \mathbf{b}, \mathbf{x} \ge \mathbf{0}$, for otherwise the problem is trivial. In the form

$$x_j \ge \alpha \sum_{i=1}^{n} x_i$$

the parameter α could be made arbitrarily large if $\mathbf{x} = \mathbf{0}$ is feasible). The optimal solution to (NLP), if one exists, provides the upper limit to the domain of $p(\alpha)$. Observe that the mere existence of a solution of $\mathbf{Ax} \le \mathbf{b}, \mathbf{x} \ge \mathbf{0}$ guarantees a feasible value of $\alpha = 0$ for (NLP) but nothing more. It is also of some practical interest that in certain cases (NLP), although feasible, will not possess an optimal solution, but this can only occur when the set of feasible solutions to $\mathbf{Ax} \le \mathbf{b}, \mathbf{x} \ge \mathbf{0}$ is unbounded.

We now show that (NLP) can be reduced to an ordinary linear program via classical methods of fractional programming, see [1]. Assume that $\mathbf{Ax} \le \mathbf{b}, \mathbf{x} \ge \mathbf{0}$ is bounded and $\mathbf{x} = \mathbf{0}$ is not feasible. Let

$$r \equiv \left(\sum_{i=1}^{n} x_i \right)^{-1}$$

and $y_j = r x_j$. After this substitution (NLP) becomes the following linear program in the variables $\{y_j\}$, r and α:

$$\begin{aligned} &\max\ \alpha \\ &\mathbf{Ay} - r\mathbf{b} \le \mathbf{0} \qquad \text{(NLP)} \\ &\sum_{i=1}^{n} y_i = 1 \\ &y_j - \alpha \ge 0,\ j = 1, 2, \ldots n \\ &y_j \ge 0. \end{aligned}$$

which may be solved directly.

ACKNOWLEDGEMENT

We wish to thank the referees for suggesting the fractional programming strategy for solving the nonlinear program (NLP) as well as certain other suggestions and improvements.

REFERENCES

1. Charnes A & Cooper WW (1962) Programming with linear fractional functionals. *Nav. Res. Logist. Q.* **5**(1). *Q.* **5**(1).
2. Hogan WW (1973) Point to set maps in mathematical programming. *SIAM Rev.* **15**(3).

AL Soyster
B Lev

(*November* 1977)

Department of Energy
Washington
D.C. 20461
USA

Temple University
School of Business Administration
Philadelphia
Pennsylvania 19122
USA

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 284–286
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0701-0284$02.00/0

Inventory Control by Budget

PRACTICAL INVENTORY CONTROL

IN A STUDY of the use of computer models in aggregate production planning, Harrison [7] has shown that although there exist such different theoretical models, few have been found to be applied in practice. He identifies a number of drawbacks related to costings and non quantifiable aspects, with which the theoretical approach is unable to cope: the real production planning environment is more complex. Eilon [5] too takes up the theme and argues that the wrong problems are being tackled, that research in some areas can bring only marginal benefit and may in any case be based on assumptions of doubtful validity.

It is perhaps natural to look upon inventory control as a subsystem of production, and in many industries this is a necessary course to take. In the more consumer oriented industries however, manufacturing may need to react more swiftly to market conditions and hence be more greatly influenced by inventory levels. Production may be seen to serve the Distribution function, rather than to dictate stock holdings based on its own criteria of effective performance. Furthermore, the reality of the situation is that each function (i.e. Production, Marketing and Distribution) is intent upon performing sufficiently well within its own respective budget, despite an awareness of the difficulties faced by the others in achieving individual and overall objectives. In consequence, the control of inventory takes on its own special character, is often the responsibility of a department which is broadly autonomous, and operates within constraints which are peculiar to each particular business.

THE THEORETICAL APPROACH

In contrast, the theory of inventory control is either allied to the theory of production control and is thereby mostly impractical, or stands as a research topic in its own right, with many examples in the literature of operational research. Seldom are claims made of implementation, as each special case is analysed and the mathematical complexities are increased. In general, contributions made to the subject of inventory control have three essential characteristics in common (also expressed by Harrison [7]): they rely upon accurate cost formulation, assume behaviour patterns which are statistically well defined and expect easy flexibility of action on the part of management.

In reality, seldom do any of these conditions pertain. Various costing methods have been proposed, and some applied in practice. Arguments can be advanced to support product costing, the allocation of distribution costs, and/or the calculation of profitability on a customer basis. As a service function, the costing of Distribution can be likened to that of Management Services and many of the issues raised by Eilon [4] on this topic are of relevance. What becomes evident, however, is that decisions taken on the basis of one method of costing can often prove different to those taken on another. Yet as Stainton [11] has indicated, from the practical standpoint, most costs of Distribution are fixed anyway.

The advantage of assuming well conditioned behaviour patterns into a research topic is that they are mathematically convenient to manipulate, but unfortunately rare to find. Nevertheless, even when a suitable pattern is properly identified, what has happened in the past is not always a predictor for the future.

MANAGEMENT INFLUENCE

More important, few managers can restrain themselves to allow events to take their course as statistical facts: their own career success or failure can depend largely upon reading signs early on and they are anxious to react and to influence the course of events by their own decisions.

Such action is frequently only possible within strictly defined departmental boundaries. Sub-optimisation occurs, described by Rivett [9] and others, which is clearly not in the interests of the firm as a whole. The overall flexibility which is implied by the researchers is a valuable ideal, but mainly impractical. Only within strict limitations may a factory stop producing: it is concerned with employment stability and labour relations which as Stainton [12] has described, are important and sometimes overriding considerations for the production scheduler on the factory floor.

There are reasons therefore to criticise the traditional approach to solving a company's inventory control problems. The solutions as they are presented are valid only in that they solve the problem as it is defined, not as it is.

DISTRIBUTION CONSTRAINTS

It is necessary therefore to look upon the control of inventory and the Distribution function as a whole in a more realistic light. In most companies, Distribution is structured the way it is because of tradition, personalities, whim, entrepreneurialism, opportunism or other such factors. Obviously, the manager responsible will want to see his function operating in as trim a fashion as he believes possible, within the constraints of the budget he has negotiated for the year to come. It is unlikely that at any given moment any of his facilities will be at an optimum setting: for example he will have too many or too few vehicles, or depot locations will not reflect the migration of population, the availability of a work force and the convenience of motorways. Similarly, customer demands will fluctuate, at least in the short-term, to such an extent that some products will be in short supply whilst others in excess. Action may be taken by Marketing to alleviate the situation, but with the inevitable lags and feedback consequences illustrated by Forrester [6].

In short, whatever the conditions, Distribution must do the best it can to operate within the constraints of its own environment, to accomodate the fluctuations,

Vol. 6
1978

to accept goods as they come and make space available as it must. In no way can the practising Distribution executive keep rigidly to the ideals of an inventory control system, the input to which he may be unable to influence effectively.

OPERATING WITHIN BUDGET

We are left therefore with the recognition that for many companies the control of inventory is about utilising existing resources of vehicles, man and depot capacity to satisfy the demands of the business within the constraints of the budget. It might be argued that the utilisation should be optimised, i.e. operate at least cost, but realities suggest that under budget achievement in one year implies the mandate of a similar objective in subsequent years, which is perhaps contrary to the concept of "maximisation of well being", proposed by Rivett [9] as one of the three major objectives of the firm.

Provided that a Distribution function can cope with its workload it does so, pruning its facilities as it can, yet aware that market conditions can change. Nevertheless, it must have certain control parameters on which its performance can be assessed. Carlson [3] proposes one way in which performance may be measured, and standards maintained, which has similarities to the procedures which have been adopted by the Distribution function of a major British food company. In this case it was possibly to identify depot capacity in the form of space availability as the major factor of control. The objective of the system is to utilise existing space in such a way that the customer service objectives can be met, within the latest production performance of the firm.

A stock shortage might be caused by insufficient production or additional demand, whilst excess stock could accrue for opposite reasons. Either way, the available stock can be distributed to the depots equitably. The Distribution function is unable to dictate what it requires, only react to circumstances in a controlled way. The question is not one of identifying accurate costs and minimising, but rather one of satisfying basic functional needs within budget.

SPACE UTILISATION

Other writers have touched upon the necessity and the reality of controlling inventory through space utilisation, for example Jones and Roper [8] and Buffa [2]. Nevertheless, each has decreed the calculation of relevant costs which in practice are hard to find. The approach which has been adopted by the food manufacturer is to achieve control within budget limits, not to attempt to identify costs across departmental boundaries which may be politically unacceptable.

The space utilisation of the company described by Stainton [13], requires a careful analysis of how much space by product in each depot is needed and more importantly, a mechanism by which it may be maintained and modified in the light of changing events. To achieve this, space in each depot is made available by product, by product group and by total production, so that if for example a particular product requires more space, it may be allowed to overflow into that space allotted to the group to which it belongs; if that space is taken up, then it may overflow into the total production space.

The rules which are applied to achieve the desired results are effective but simple. Given a delivery fleet of vehicles which operate regularly from factories to depots, a prime consideration is that they should be fully loaded. In addition, there is limited storage space at the factories and so whatever is produced above predetermined levels, is required to be moved out and distributed to the depots. Any system which relied only upon depot forecasts of demand would not necessarily clear all of the stocks from the factories: equally, any depot which over-reacted to a rising trend (or had inside knowledge about an impending shortage), could easily take up more stock of the relevant food lines than it should, thus causing an artificial need for further production.

DEPOT ALLOCATION

The answer lies in allocating stocks to the depots from the factories on an equitable basis, which depends upon the characteristics of the products themselves. Fortunately a significantly large number of the company's products have production capacities which match demand. For these products it is in the interests of Distribution to deliver regular fixed quantities to the depots.

To illustrate, let the fixed quantity of product i to depot j be q_{ij}. Suppose also that f_{ij} is the forecast for product i in depot j, S_{ij} is the safety stock required, a_{ij} is the current level of stock and w_{ij} is the amount of stock which would take up the space allocated to the product i in the depot j. Then it is required that

$$f_{ij} + S_{ij} \leqslant a_{ij} + q_{ij} \leqslant w_{ij} \text{ for all } i \text{ and } j, \ldots. \quad (1)$$

whilst $\Sigma_i\, q_{ij} \leqslant Q_j$ for all j, where Q_j is the quantity of product j available for distribution.

If inequality (1) cannot be satisfied, then it is required that $\Sigma_i\,(a_{ij} + q_{ij} - w_{ij}) \leq L_{jk}$ for all j and k, where L_{jk} is the space allotted to product group k in depot j. The quantities q_{ij} are sent from the factories if these constraints are satisfied. If they are not, then management are able to take other pre-arranged action, such as to send the goods to another depot which has space, to retain the stock in the factory, to resort to outside contractor warehousing, and ultimately to seek a slow down in production.

Clearly, some products have a steadier demand than others. Much variation can be contained within the system, particularly if the depots are large enough to be able to afford a generous share of space for each product. Even so, the products with considerable seasonal fluctuations will require modifications to the depot spaces allocated to them. In practice, this occurs 3 or 4 times a year. This action is taken collectively by a management group representing the interests of Marketing, Distribution and Production. They are also concerned with any other matters, such as exceeding space, for which the system triggers management intervention. Their involvement, including day to day affairs such as the monitoring of sales promotions, correcting delivery errors and occasional staff shortages amounts to the equivalent of two full time people; it has considerable advantage by ensuring fast communication between the interested parties.

BATCH PRODUCTS

Some products are manufactured at a far greater rate than the demand for them. For these, it is not possible to employ directly the constraints described above. Instead, the work of Simpson [10] and Brown [1] have been found to be helpful.

Not only should the batch products (as we shall call

them) be distributed equitably to the depots, but it is far more important to avoid overstocking at any one depot and shortages at others since additional production might be difficult to arrange at relatively short notice, and the company does not wish to redistribute. It is appropriate to make deliveries of the batch products at regular but more extended intervals, say every two or three weeks, for which purpose the delivery fleet had been designed. When allocating batch products to depots, therefore, it was desirable to target for all depots to run out of stock simultaneously over a period greater than the interval between deliveries, taking full account of stocks left in the depots at the time of allocation as well as new stock that had been made recently, or kept back from a previous production run at the factory.

The equations may be devised as follows:

Assume depot i is about to run out of stock, but receives delivery Δ_q. This small quantity can be expected to last for Δ_q/x_i, where x_i is the expected demand from depot i. During that time, other stocks will have been depleted by $(X - x_i)\,\Delta_q/x_i$ where X is the total expected demand over all depots.

Let p_i be the probability that depot i will be the first to run out of stock. Then the change in total stock is

$$\Delta R = -(X - x_i)\frac{\Delta q}{x_i}p_i + \Delta q(1 - p_i).$$

so

$$\frac{dR}{dq} = 1 - p_i\frac{X}{x_i}$$

We would wish to arrange for all depots to run out of stock at the same time, which is equivalent to seeking that remaining total stock be at a minimum when the first depot runs out. In consequence, we require that $p_i = x_i/X$.

Let $\alpha\,(t)$ be the probability that any one of the depots will run out of stock before time t has expired. Then the probability that depot i will be the first to run out before time t is $\alpha(t)\,p_i$, which defines a safety factor k_i.

Suppose that the first depot runs out of stock at time t. The planned remaining stock at this time is

$$k\sigma(t) = (0.659 + 0.341t)\sum_i k_i\sigma_i$$

where σ_i is the standard deviation of forecast errors at depot i over one unit of time. Now

$$xt + k\sigma(t) = Q + \sum_i s_i$$

where Q is the total stock available for allocation from the factory, and s_i is the remaining stock at depot i. Then

$$\sum_i k_i\sigma_i = \frac{Q + \sum_i s_i - Xt}{(0.659 + 0.341t)}.$$

Each of the variables on the right hand side is known (t takes the value 2 or 3), as are the values of the σ_i. If we now assume that the forecast errors are distributed normally, then we can identify the k_i in association with the αp_i, which is equal to $\alpha x_i/X$. In effect, we are seeking values of the k_i which satisfy the above equation, whilst recognising that the probabilities from which they are derived are directly proportional to the fraction of total demand in each respective depot.

Once the k_i have been determined the quantity to be sent to the ith depot is $q_i = x_i t + (0.659 + 0341t)\,k_i\sigma_i - s_i$.

If q_i turns out to be negative, that depot is excluded from the analysis and the procedure is repeated without the ith depot. Factory stock is treated as though it is at a depot.

CONCLUSION

In attempting to find a system solution to the food manufacturer's stock imbalance problems, it was soon discovered that the classical methods of inventory control were impractical, mainly because it proved impossible to identify relevant costs. The reality of the situation was that the company functioned within the constraints of budget, and that existing facilities had to be used to best advantage. It is contended that such a situation is the rule rather than the exception. The system, described in broad outline above, has been operational since the early 70s.

REFERENCES

1. Brown RG (1967) Decision rules for inventory management. Holt, Rinehart and Winston, New York.
2. Buffa FP (1977) A model for allocating limited resources when making safety stock decisions. *Dec. Sci.* **8**, 415–426.
3. Carlson (1977) Maintaining the calibration of an information system. *Omega* **5**, 271–280.
4. Eilon S (1976) Editorial—how to charge for management services. *Omega* **4**, 623–628.
5. Eilon S (1977) Editorial—who says production control is dead? *Omega* **5**, 107–113.
6. Forrester JW (1961) *Industrial Dynamics.* MIT Press, Cambridge, Massachusetts.
7. Harrison FL (1976) Production planning in practice. *Omega* **4**, 447–454.
8. Jones WG and Roper CM (1964) Linear programming applied to production planning—A Case Study. *Opl Res. Q.* **15**, 293–302.
9. Rivett BHP (1972) *Principles of Model Building.* Wiley, London.
10. Simpson KF Jr (1958) A theory of allocation of stocks to warehouses. *Opns Res.* **6**, 863–873.
11. Stainton RS (1976) Fixed cost constraints and realities. *Omega* **4**, 747–748.
12. Stainton RS (1977) Production scheduling with multiple criteria objectives. *Opl Res. Q.* **28**, 285–292.
13. Stainton RS (1977) Practical space constraint restrictions. *Scandinavian J. Mater. Admin.* **2**, 64–69.

Dr RS Stainton

Henley
The Administrative Staff College
Greenlands
Henley-on-Thames
Oxfordshire RG9 3AU
UK

(*November* 1977)

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 3, pp. 287–289
© Pergamon Press Ltd 1978. Printed in Great Britain

0305-0483/78/0701-0287$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the Lyon Playfair Library, Imperial College of Science and Technology, Exhibition Road, London SW7 2BX, UK:

ABRAHAM RJ (1976) Markov Chain Models of Share Prices. MSc Thesis, Imperial College, London.

ALLEN JA (1976) An Interactive Version of GASP. MSc Thesis, Imperial College, London.

BAKER RC (1976) The Application of a System Dynamics Model to the Production–Distribution System for Pre-Recorded Tape Products. MSc Thesis, Imperial College, London.

BARTON CP (1976) Worker Participation and Industrial Democracy. MSc Thesis, Imperial College, London.

BHANJI YKH (1976) Financial Planning Models for a Manufacturing Company. MSc Thesis, Imperial College, London.

BRITTEN N (1976) A Graph Theory Approach to International Trade. MSc Thesis, Imperial College, London.

CARDOSO T (1976) An Application of AID (Automatic Interaction Detector) to a Study of Wastage. MSc Thesis, Imperial College, London.

CAVE BJ (1976) Unemployment, Inflation and the Labour Market. MSc Thesis, Imperial College, London.

CHEONG PF (1976) A Study of the Profitability Levels of Departments in a Department Store. MSc Thesis, Imperial College, London.

CLARK–WILSON DA (1976) Property Bonds. MSc Thesis, Imperial College, London.

COBBIN SJ (1976) Aspects of the Design of a Computerised Company Cost Model. MSc Thesis, Imperial College, London.

DIXON P (1976) Financial Modelling via Systems Dynamics. MSc Thesis, Imperial College, London.

EDWARDS TA (1976) Analysis of Post Office Telecommunications Market Research Data Using Automatic Interaction Detector. MSc Thesis, Imperial College, London.

FORREST BA (1976) An Econometric Model of the Vehicle Refinishing Paint Market. MSc Thesis, Imperial College, London.

FRENCH DL (1976) A Case Study in the Footwear Industry (The Effects of a Change in the Pattern of Trade on the Relationship between a Manufacturer and his Customer. MSc Thesis, Imperial College, London.

HERVITZ HM (1976) Resource Allocation in Decentralised Organisations. MSc Thesis, Imperial College, London.

HORE F (1976) The Organisation of Petroleum Exporting Countries, a Study into the Nature of the Oil Cartel. MSc Thesis, Imperial College, London.

HOWELL IV (1976) Developments in Ethylene and its Major Derivatives. MSc Thesis, Imperial College, London.

KALIRAY JS (1976) Strikes. MSc Thesis, Imperial College, London.

KAVANAGH RT (1976) Short Term Forecasting of Electricity Sales Using Leading Indicators. MSc Thesis, Imperial College, London.

KWAN TQ (1976) Pricing and the Economic Structure of the World Aluminium Industry. MSc Thesis, Imperial College, London.

KOH SC (1976) Terminal-Based Enquiry Systems. MSc Thesis, Imperial College, London.

LAMBERT GR (1976) Operating Costs of a Transport Fleet. MSc Thesis, Imperial College, London.

LAWLESS MF (1976) An Analysis of South African Agricultural Exports to the European Economic Community with Particular Reference to Tariffs and Other Barriers to Trade. MSc Thesis, Imperial College, London.

LEHMANN NR (1976) The Foreign Exchange Role of a Treasurer in a Multinational Company. MSc Thesis, Imperial College, London.

LEONIDIS–PLESSAS D (1976) Passenger Air Traffic Dynamics in Air Transport. MSc Thesis, Imperial College, London.

LEVIN D (1976) Aspects of Sensitivity Analysis for Arbitrage Operations in Foreign Exchange. MSc Thesis, Imperial College, London.

LISMER PA (1976) International Taxation. MSc Thesis, Imperial College, London.

LOVETT DR (1976) The Utility of Further and Higher Education. MSc Thesis, Imperial College, London.

LUCAS PBB (1976) The Market for Fine Table Wine in the United Kingdom. MSc Thesis, Imperial College, London.

MAYS IA (1976) Cash Flow Forecasting in a Service Industry. MSc Thesis, Imperial College, London.

MCCULLAGH AM (1976) Graph Theory and Linear Programming to Alleviate Exchange Risk in a Multinational Company. MSc Thesis, Imperial College, London.

MISIRLIS GI (1976) The Vehicle Scheduling Problem. MSc Thesis, Imperial College, London.

MOORE AF (1976) Forecasting Using Leading Indicators. MSc Thesis, Imperial College, London.

MORGAN SL (1976) A Review of the Capital Asset Pricing Model and its Application to Company Investment. MSc Thesis, Imperial College, London.

NICHOLAS PK (1976) Wage Rounds and Key Bargains. MSc Thesis, Imperial College, London.

NICHOLLS LC (1976) Some Techniques for the Scheduling of Dredgers. MSc Thesis, Imperial College, London.

PIOTROWSKI AM (1976) Managing Foreign Exchange Operations: Some Topics. MSc Thesis, Imperial College, London.

POOLE SF (1976) An Exercise in the Application of Graph Theory to International Tax Planning. MSc Thesis, Imperial College, London.

PRICE (1976) Incomes Policy in the United Kingdom—An Assessment. MSc Thesis, Imperial College, London.

RANDALL GE (1976) Tendering and Contract Implementation. MSc Thesis, Imperial College, London.

RECORDATI GM (1976) The Decision Making Process in Research and Development. MSc Thesis, Imperial College, London.

ROSEN HJ (1976) A Study of Pricing Policy in the United Kingdom. MSc Thesis, Imperial College, London.

RUDOLF F (1976) Estimating Market Share Functions for a Branded Consumer Product. MSc Thesis, Imperial College, London.

SELLMEYER D (1976) An Alternative Heat Source—A Comparison of Domestic Solar Water Heating Systems in the United Kingdom. MSc Thesis, Imperial College, London.

SELLMEYER W (1976) Computer Programs for Queueing Models. MSc Thesis, Imperial College, London.

SENDRA L (1976) Aspects of Participation in Chile. MSc Thesis, Imperial College, London.

SINGIAN VI (1976) The Effects of Currency Developments—A Programme for Arbitrage. MSc Thesis, Imperial College, London.

SONDHI P (1976) Some Aspects of Investment Appraisal in the North Sea. MSc Thesis, Imperial College, London.

STATHER LJ (1976) Forecasting of Economic Activity Using Bank Clearing Statistics as a Leading Indicator. MSc Thesis, Imperial College, London.

SUTHERLAND W (1976) Attitudes to Economic Planning in the United Kingdom and France. MSc Thesis, Imperial College, London.

TING YS (1976) Formulation of a Leasing Model for a Financial Leasing Company. MSc Thesis, Imperial College, London.

TSIMBOUKIS D (1976) Seasonal Inventories and Aggregate Production Planning in a Paint Manufacturing Firm. MSc Thesis, Imperial College, London.

TUAKLI D (1976) The Decision Making Process in the Evaluation and Selection of Research and Development Projects. MSc Thesis, Imperial College, London.

WILLIAMS MR (1976) An Analysis of the Financial Needs of an Industry: Electronic Computers. MSc Thesis, Imperial College, London.

WU NK (1976) Production Control—Information Recording and Processing. MSc Thesis, Imperial College, London.

The following are obtainable at the University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, USA:

BOURNE LD (1976) Order Effects in Accounting Variance Reports: a Laboratory Experiment. Doctoral Thesis, Washington University, MO.

BROCKHAUS RH (1976) Locus of Control and Risk-Taking Propensity as Entrepreneurial Characteristics: a Comparative Study. Doctoral Thesis, Washington University, MO.

HALLORAN JA (1976) The Effect of Mortgage Form on Borrower and Lender Risk. Doctoral Thesis, Washington University, MO.

HIRSCH ML (1977) Disaggregated Probabilistic Accounting Information: an Experiment to Discover Systematic Dysfunctional Behavior. Doctoral Thesis, Washington University, MO.

HOOVER JD (1974) Experimental Learning Effects on Emphatic Interpersonal Communication Skills and Related Attitudes. Doctoral Thesis, Washington University, MO.

HUMPHREYS MAH (1977) A Comparison of Some Data Collection Methods for Nonmetric Multidimensional Scaling. Doctoral Thesis, Washington University, MO.

KIM KY (1975) Planning a Multi-Echelon Production–Distribution System in the Petroleum Industry with a Simulation Model. Doctoral Thesis, Washington University, MO.

MARSHALL WJ (1977) A Normative Model of Bank Portfolio Selection under an Equity Value Criterion. Doctoral Thesis, Washington University, MO.

MESSMER DJ (1974) Belief Discrepancy and Repetition as Determinates of the Affective Response to Television Advertising. Doctoral Thesis, Washington University, MO.

PATTON JM (1975) Usefulness of Municipal Financial Reports. Doctoral Thesis, Washington University, MO.

SEATON FB (1974) The Impact of Objective Information on Consumer Brand Preferences: an Experimental Investigation. Doctoral Thesis, Washington University, MO.

Taylor MR (1974) Consumer Expectations, Product Information Value and Decision Commitment: Effects on Consumer Product Satisfaction. Doctoral Thesis, Washington University, MO.

Woerheide WJ (1977) The Effect of Quarterly Earning Announcements on Estimates of Systematic Risk. Doctoral Thesis, Washington University, MO.

Wright TV (1976) A Comparative Performance Analysis of Black-Owned Commercial Banks and White-Owned Commercial Banks Located in Black Communities in Unit Banking States in 1973. Doctoral Thesis, Washington University, MO.

Yeung ETM (1974) Environmental Uncertainty and Organizational Designs. Doctoral Thesis, Washington University, MO.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 6
1978

BLANK
PAGE

Vol
197

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 291–293
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0291$02.00/0

Vol. 6
1978

EDITORIAL

How to Analyse Changes in Performance

ANALYSES of performance of business enterprises are often confined to statements about profit and loss accounts, to comparisons of key figures in one year with the corresponding figures in the previous year, and to computations of certain selected ratios regarded as major indicators of how well or otherwise the enterprises in question have been operating.

The danger of looking at any performance criteria in isolation is well appreciated by experienced analysts. And yet the simplicity of these criteria has such a mesmerising effect that people tend not only to accept them at face value and to become ignorant of the significant way in which criteria are inter-related, but also to become unaware of the root causes that lead to changes in performance.

Take, for example, the much misunderstood yet widely quoted term 'productivity'. It lies at the heart of any assessment of industrial effectiveness, and it is used of course in many inter-firm and inter-country comparisons. To most people 'productivity' simply means the ratio of output per man (or man–hour), but such a measure taken in isolation has the implication that labour is the only input to an industrial enterprise, whereas other inputs (such as materials and capital) have in fact their own contributions to make. Thus, capital expenditure aimed at increasing capacity or reducing the labour content per unit of output may be crucial if the general performance of a firm is to improve, and what in a particular case seems at first sight a significant increase in labour productivity may turn out at closer examination to be largely due to the productivity of capital. Similarly, an increase in output per man may be completely offset by the extra cost of capital or materials involved, so that taken on its own the labour productivity index may increase while the profitability of the enterprise may simultaneously decline.

The problem is further exacerbated by the lack of consistent and coherent definitions. What, for example, is precisely meant by 'output'? If a firm produces a single product, the specifications of which remain unaltered over a period of time, output is easily measurable by the number of physical units produced. But in multi-product systems, and when the product-mix changes, physical quantities become meaningless. The remedy that is often resorted to is a measure of output in terms of its sales value, but then it is possible for such output per man to increase simply because of price increases, even when physical output actually goes down.

Take another example, which concerns that widely accepted criterion of performance, the rate of return. In its simplest form this measure may be expressed as follows:

$$\text{Rate of return} = \frac{(\text{unit price} - \text{unit cost})\ \text{volume}}{\text{total investment}}$$

Now, what would be the consequences of a change in each of the parameters on the right hand side of this equation? The important point to realise is that a primary change in one parameter causes a series of secondary changes in others. For example, if unit price (namely the average price charged per unit of output) increases, volume may well decline owing to the effect of price elasticity of demand. This reduction in volume must lead to an increase in unit cost, simply because the fixed cost element in the total cost structure is then distributed among a smaller production output. Thus, even if total investment were to remain unchanged, it is not immediately clear what the total effect of a price increase would be. It is only when one can ascertain first that the increase in unit price will outweigh the increase in unit cost, and secondly that the reduction in volume will be more than compensated for by the increase in profit margin, that an increase in the rate of return can be expected.

Further complications arise when other factors are brought in. For example, if the cost element associated with interest charges is separately identified in the unit cost equation, then this element is also affected by the production volume; the total interest charge is a function of the total borrowing which is affected by the working capital, and this in turn may change as the total revenue or the total cost changes as a consequence of a change in price or volume or both. And, needless to say, a change in working capital leads to a change in total investment and thereby has a further influence on the rate of return.

We see, therefore, that the major variables that determine the level of performance (the comments made earlier are essentially valid for performance criteria other than the rate of return) are connected in an intricate network of relationships, and a change in any one variable triggers off a series of perturbations in the system, some of which may even have an effect on the primal cause itself. And it is only through the scanning of the whole network, and after taking account of all the possible secondary effects, that the final outcome with respect to given performance criteria can be properly ascertained.

The computational work involved in any *ad hoc* case is obviously a tiresome affair, but that aspect of the problem is the least worrying from the analyst's viewpoint. With the computing power that is readily available today it is neither difficult nor very expensive to run a large number of scenarios, each reflecting a change in one of the parameters or a set of simultaneous changes in several parameters. The danger is that the analyst, and with him the manager, will not see the wood for the trees. What is important for them both to understand is the underlying pattern of cause and effect, coupled with sensitivity ana-

lyses, so that the major variables in the system can be singled out for closer scrutiny and control.

This is where a method for incremental analysis can be so useful. As the name implies, such a method involves a model of behaviour of relative incremental changes, as for example the change in unit price relative to the original price, the change in volume relative to the original volume, and so on. In concentrating on relative increments rather than on absolute values it is much easier to construct the network of relationships mentioned earlier and to explicitly account for them in the analysis. It is also possible to decompose such a model to any level that is thought to be appropriate for a given situation, for example by separating fixed or variable costs into several cost items, by differentiating between several sources of funds borrowed at different rates to finance working capital, by disaggregating the output into separate product groups, and so on. Similarly, the approach is applicable to various levels of aggregation in a given organisation, and the analysis can be employed both for an enterprise as a whole as well as for selected parts of it.

The method of incremental analysis allows not only a systematic investigation to be carried out on past performance, but also provides a means to assess the possible effects of future changes, such as wage increases, changes in the cost of materials, capital expenditure, interest rates, levels of demand, and so on. Thus, the method can become an invaluable planning tool in managerial decision making.

Vol. 6
1978

SAMUEL EILON
Chief Editor

BLANK
PAGE

Vol.
197

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 4. pp. 295 304
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0295$02.00/0

Productivity Drift in Extended Learning Curves

DR TOWILL

U KALOO

University of Wales Institute of Science and Technology, Cardiff, UK

(*Received January* 1978)

The pattern of changes in the Extended Learning Curves representing the performance of 114 assembly workers in an electronics firm is examined in terms of changes in productivity drift over long periods of time. It is shown that 2 types of discontinuities can occur. The first consists of a temporary levelling off in productivity followed by a recovery to new higher levels of productivity; the second has a sustained period of negative productivity drift. A number of work facets are statistically tested as possible causes of these patterns of productivity drift, but, the only significant relationship (at $P \leqslant 0.05$ level) is with length of service in job. Some evidence is presented to show that conventional work measurement methods are not effective in terms of assessing the true work output of experienced workers since such workers are able to 'beat the system' by claiming those allowances which are most difficult to cross-check.

Vol. 6
1978

1. INTRODUCTION

IT IS well known that the cycle time to perform industrial tasks decreases with experience or time on the job. This is termed "learning" and the first recorded observations are probably the studies of telegraphy by Bryan and Harter 1897 [19]. Plots of performance measured during the "learning" or "startup" phase, are referred to as "learning curves". Various axes may be used in the construction of learning curves [2], and equations describing the curves are known as learning curve models, a wide variety having been already proposed in the literature, for example, de Jong [7], Wright [23], Glover [8, 9], and Bohlen and Barany [3]. Using the output per unit time as the dependent variable, and time on the job as the independent variable of the learning curve has a number of advantages when the curve is used as a management tool [21], and naturally leads to the use of a particular learning curve model defined by,

$$Y(t) = Y_c + Y_f(1 - e^{-t/\tau}) \qquad (1)$$

where $Y(t)$ is the productivity level at time t, Y_c is the initial productivity level, and Y_f is the ultimate increase in productivity level. Equation (1) is referred to as the "time constant learning curve model", and is designed to model performance during the "startup" or "learning" phase which for the purpose of this paper may be arbitrarily defined as terminating at $t = 3\tau$, where τ is the learning time constant [2, 12]. Since τ has been shown to vary from three weeks to twenty-four weeks for a variety of industrial tasks [11] the calendar duration of a study to observe learning phenomena may range from two months to two years. This paper is concerned with the extended learning curve, as shown in Fig. 1, in which two regions are identified as follows,

(a) "learning" or "startup" phase for $t \leqslant 3\tau$

(b) "productivity drift" phase, for $t > 3\tau$.

Beyond $t = 3\tau$, although equation (1) has an ever decreasing positive slope, which is not unreasonable for drift performance, it should not be expected to model performance particularly during this phase, as will be seen later. This study is concerned with time-dependent changes in individual productivity in the drift phase.

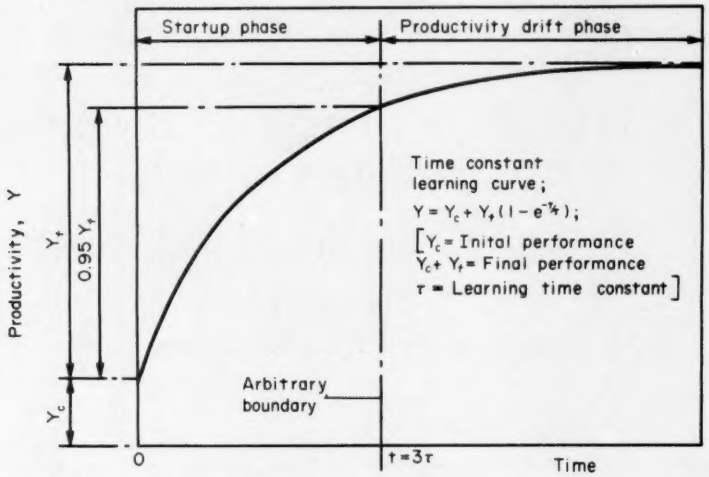


FIG. 1. *Time constant learning curve model.*

2. PRODUCTIVITY DRIFT

Previous studies suggest that the extended learning curve can take one of 3 forms in the drift phase as shown in Fig. 2.

(a) *Model I describes continued improvement in output* and could be the time constant model extrapolated for $t \gg 3\tau$. This characteristic feature may be attributed to continued learning over very long periods of time as shown in the empirical findings of Bevis [2] and Murrell [15]. It should be noted that although the production rate increases for all time, the rate of change in production rate rapidly becomes small, and performance improvement then needs special tools for detection [22].

(b) *Model II allows a definite ceiling* for the effect due to such causes as group norms and pressures in operation [5, 16, 17, 18]; and/or the effects of externally imposed limitations such as machine pacing [1].

(c) *Model III further allows for a definite deterioration in performance* due to changes in individual work behaviour [13].

Irrespective of the true form of the learning curve in the drift phase, it may be noted that because of the characteristically small slopes,

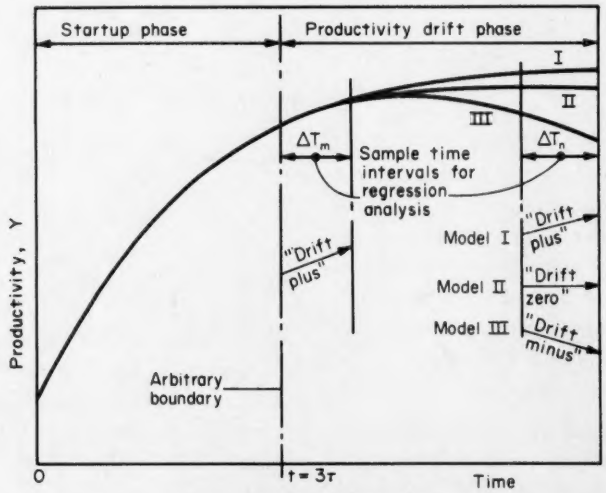


FIG. 2. *Productivity drift patterns.*

over short intervals of time, the time series data may be adequately described with a linear model:

$$Y = A + Bt \quad (2)$$

where A, B are model parameters, t = measure of time, Y = measure of productivity.

Because of the small differences in the values of Y which are encountered in the drift phase, in combination with the short term fluctuations in individual performance, productivity changes are more meaningfully expressed in terms of the change in the slope of the linear regression estimates over short intervals of time [12].

(a) Positive productivity drift (drift plus)—slope of equation (2) is significantly positive.

(b) Negative productivity drift (drift minus)—slope of equation (2) is significantly negative.

(c) Zero productivity drift (drift zero)—slope of equation (2) is not significantly positive or significantly negative.

In terms of drift patterns model I may be described as a sequence of drift plus followed by drift zero; model II may be described with drift zero for $t \gg 3\tau$; and model III will include all three drift patterns of Fig. 2.

3. OBJECTIVES OF THE PRESENT STUDY

This study is designed to investigate the pattern of productivity drift in the extended learning curve—i.e. the pattern of distribution of drift plus, drift minus and drift zero, and investigate some possible causes of particular patterns of productivity change including the explanation of large claims of "excess time" by experienced workers. There is no reason to suppose that industrial operators will always produce productivity patterns which can be described entirely in terms of Fig. 2 and this will become evident later in the paper when the discontinuity patterns are considered. All productivity figures used in the study are available from bonus records, no direct work measurement being permitted.

4. SETTING OF STUDY

This study was carried out in the switch assembling division of a large electronics firm situated in the United Kingdom. Total employment in the firm included about 900 manual workers. The firm provides a major source of employment to the population of the small town in which it is set. In the switch assembling division, batch production is used and assembly line techniques are employed in a few sequences. All the workers are women and at the time of the study, had been with the firm for periods ranging from 25 years to 1 year. The age range was 17 years to 57 years. At the start of the study all the workers had already attained the basic times set by the Work Study department and thus in the absence of any other evidence may be assumed to be in the drift phase of their individual productivity curves.

4.1. *Payment system*

The payment system consists of a fixed basic wage plus a variable bonus based on a 'rated performance level' (Y_1). The basic wage is dependent upon the length of service of the worker and the mean weekly wage for the test group at the time of the study was approximately £25.00. The bonus scheme defines 9 discrete levels of Y_1 and corresponding bonus rates. These bonus rates vary from a minimum of £1.50 per 40-hr week for $Y_1 = 115$, to a maximum of £4.00 per week for $Y_1 = 165$. It may be noted from Fig. 3 that the mean rated performance for the whole division has been substantially stable within a range of 10 units of Y_1 from week 70 to week 152 suggesting that bonus payments have been substantially stable for the past $1\frac{1}{2}$ years. Workers may move up or down the Y_1 scale by maintaining an improved or diminished level for 3 consecutive weeks.

4.2. *Productivity measure*

Productivity is here defined as the efficiency of production, i.e.

$$\text{productivity } (Y) = \frac{\text{Actual Output Rate}}{\text{Expected Output Rate}}$$

which for batch production is conveniently expressed

$$Y = t_2/t_1 \quad (3)$$

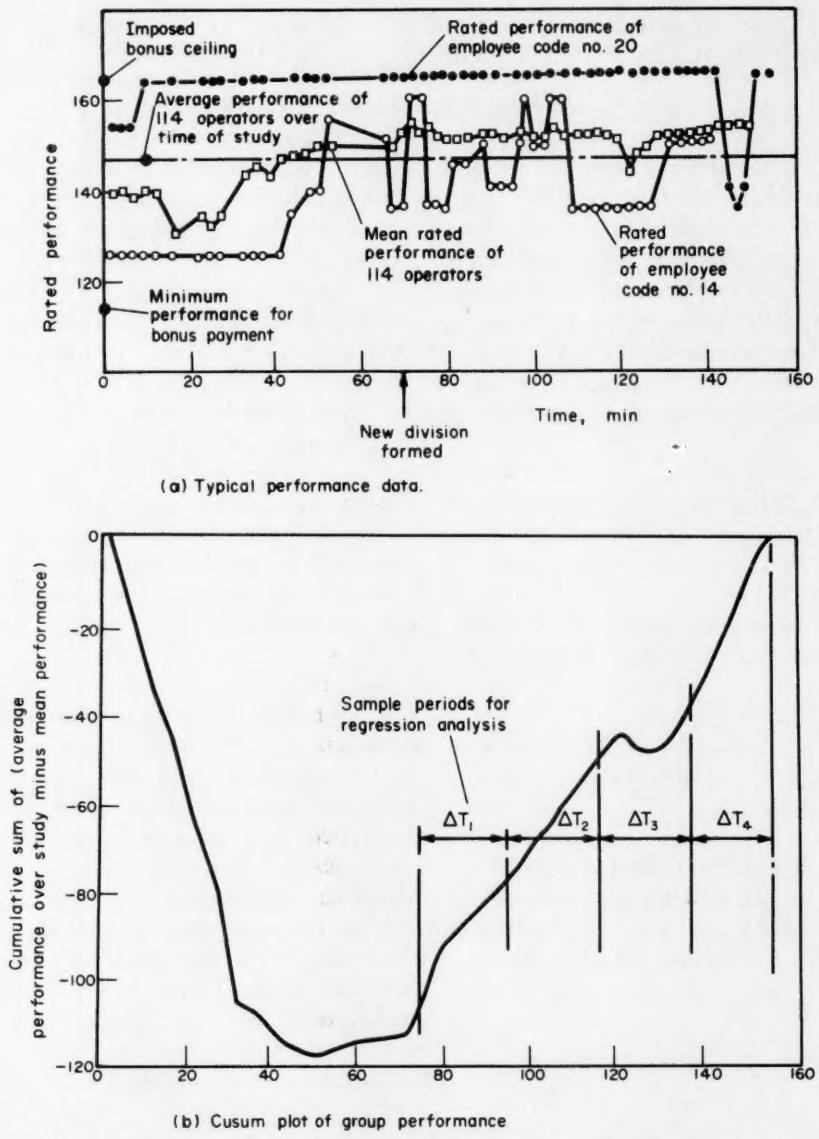

Imposed bonus ceiling
Rated performance of employee code no. 20
Average performance of 114 operators over time of study
Mean rated performance of 114 operators
Rated performance of employee code no. 14
Minimum performance for bonus payment
Rated performance
Time, min
New division formed
(a) Typical performance data.
Cumulative sum of (average performance over study minus mean performance)
Sample periods for regression analysis
ΔT1
ΔT2
ΔT3
ΔT4
(b) Cusum plot of group performance

FIG. 3. *Group and samples of individual performance. (a) Typical performance data. (b) Cusum plot of group performance.*

where t_2 = expected time or allowed time, t_1 = actual time = (Total attendance hours) − (Unmeasured Work).

'Unmeasured Work' is time spent under various headings, for example allowances for union meetings and allowances for excess work caused by poor quality parts, during which period output is not rated for purposes of bonus payments. During such periods, ratings achieved at a previous period are applied. Where unmeasured work exceeds 75% of weekly attendance, the whole week is classified as unmeasured, thus having no effect on bonus gradings. Unmeasured work claims are classified under 32 Excess Codes and 10 Delay Codes. The Excess Codes refer to claims for excess work which may be attributed to features of the work which are beyond the control of individual workers—for example, excess work caused by defective assemblies, and close limits on piece parts; the Delay Codes refer to enforced work stoppages such as machine breakdowns and personal needs such as 'visit

to surgery'. On average, claims under Excess Codes account for approximately 80% of total unmeasured work claimed by the division.

Y calculated from equation (3) above is rounded off to the nearest multiple of 5 to determine Y_1 which forms the basis for bonus payments described in section 4.1. Rated performance for the group of operators, and some sample operators, are shown in Fig. 3. The cusum plot is also shown, having been found to have useful "smoothing" properties for viewing drift phase industrial data [22].

As a measure of productivity, Y_1 has 3 important properties:

(a) There is an imposed ceiling on individual productivity at $Y_1 = 165$. The productivity curve for employee No. 20 in Fig. 3 illustrates this ceiling. Also shown for comparison are results for employee No. 14, who does improve initially, and the average performance of the whole shop over the 152 week period.

(b) The '3-up, 3-down' criterion for changes in Y_1 serves as a moving point average and damps out the short-term fluctuations in individual performance (see [12]). This effect is illustrated for the case of operator No. 75 in Fig. 4(a).

(c) The inclusion of 'unmeasured work' in the productivity computation can provide workers with a means of exerting some control over their own productivity ratings independent of individual work effort.

5. EXPERIMENTAL DESIGN

The raw data used in this study includes individual productivity data covering a period of 152 weeks, detailed individual daily work records covering a period of 3 months and service data from personnel records.

Although individual productivity data covering a period of 152 weeks was available, this study analyses the patterns of productivity drift change in the period week 70 to week 152. This is because in week 70 a new division was formed to produce one line of switches and has continued in this particular line of product since. Thus over this period of time, details of work setting such as co-workers and job description have remained substantially unchanged. The procedure adopted here is:

(a) To divide the period, week 70 to week 152, into a suitable number of short intervals of time (ΔT) over which to determine individual productivity drift via equation (2).

(b) To examine the pattern of productivity drift changes.

(c) If difference in drift pattern exist, to examine differences in work design and work structure factors which can possibly account for these differences.

5.1. *Determination of* ΔT

The choice of ΔT was based on the following considerations:

(a) It is desirable to compute drift using as large an interval, ΔT, as possible, partly to increase the statistical validity of the computed result, and partly to reduce the effects of errors introduced by the short-term fluctuations in individual productivity.

(b) Preliminary scatter diagrams of individual productivity data showed that productivity changes remained substantially linear for periods of up to 6 months. For the case of the mean divisional productivity, this is seen in the cusum plot in Fig. 3(b). The overall mean (mean over 152 weeks) was used as the target for the cusum plot.

(c) The test period included 2 vacation closedowns per year (in August and in December). To avoid discontinuities in work during a particular ΔT interval, it was necessary to arrange these breaks to fall between successive intervals of ΔT.

Thus individual productivity drift was computed over 4 discrete time intervals ΔT_1 to ΔT_4 where $\Delta T_1 = 20$ weeks, $\Delta T_2 = 22$ weeks, $\Delta T_3 = 18$ weeks and $\Delta T_4 = 16$ weeks. These intervals are superimposed on Fig. 3(b).

5.2. *Patterns of productivity drift*

To determine drift, conventional least squares linear regression analysis was applied to the blocks of productivity data in the time intervals, ΔT_1 to ΔT_4, to determine the model

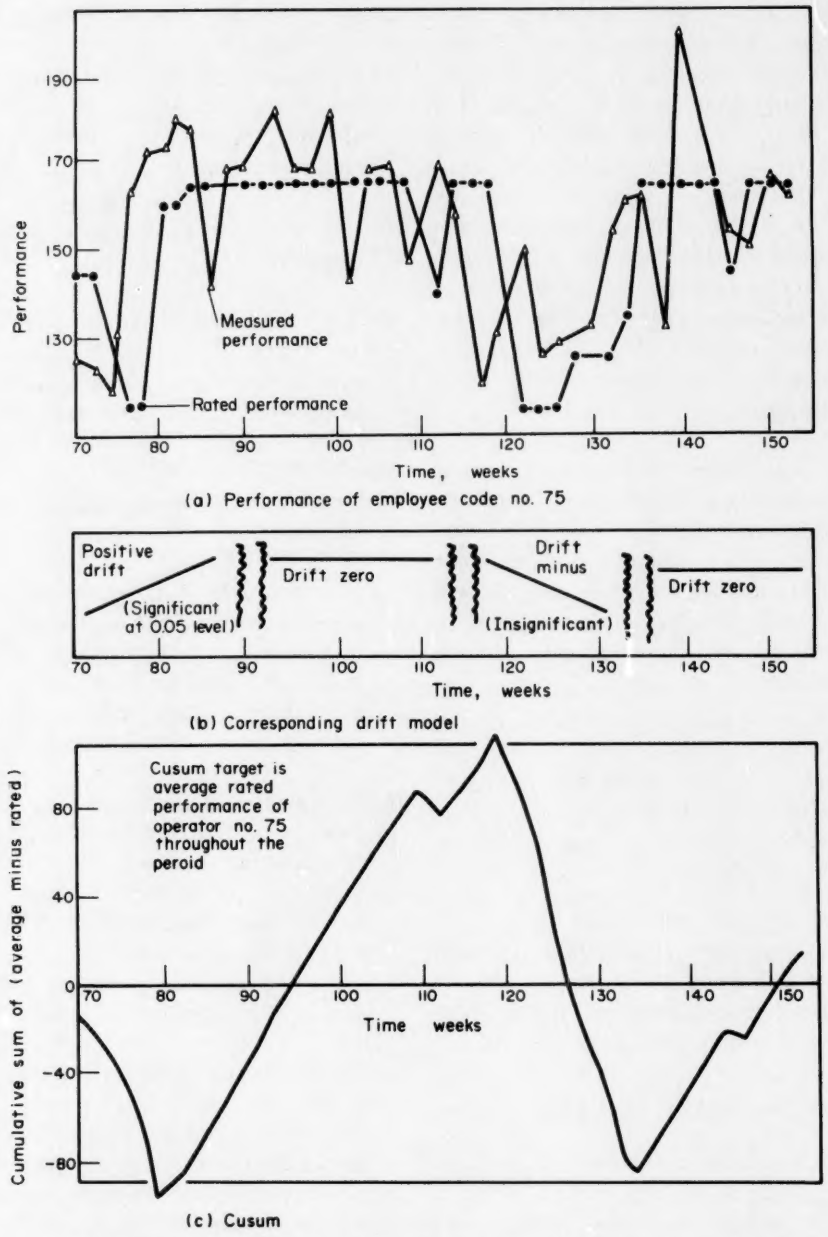


FIG. 4. *Drift performance of an individual employee.* (*a*) *Performance of employee code no. 75.* (*b*) *Corresponding drift model.* (*c*) *Cusum.*

parameters A and B in equation (2). The sign of B provided the basis for attaching one of the three values that have been assigned to drift in section 2. Productivity changes expressed in terms of productivity drift, as observed in the 'B' coefficient are illustrated for operator No. 75 in Fig. 4(b). For this particular operator, there is only one slope significant at the 0.05 level. This is "drift positive" and occurs during the first interval ΔT_1.

5.3. *Possible explanations for similarities and differences in drift patterns among individuals.*

In any industrial setting a very large number of environmental or external factors can effect the measured productivity of individual

workers, e.g. availability of raw material, running down of machines and tools, union policy changes, supervisory style, company policy changes, and so on. The following hypotheses were set up to test if the measured differences in individual productivity drift could be attributed to any of these more obvious assignable causes.

Hypotheses 1. (*H*1) *Productivity drift is dependent upon* sub-group membership *within the division.* There are 10 sub-groups within the division each specialising in one product or process. There is some transfer of operators between sub-groups but this seems to be on a temporary basis not usually exceeding one day. Over the test period the sub-group membership remained substantially the same. This hypothesis anticipates the following,

(a) the 'atmosphere' in certain sub-groups is more conducive to work.

(b) certain sub-groups may consistently get easier tasks.

(c) certain sub-groups may have developed methods of 'beating the system'.

(d) particular sub-groups may suffer selectively from frequent machine breakdowns or lack of components.

*H*2. *Productivity drift is dependent upon supervisor.* Seven of the smaller groups come under one supervisor and the three larger groups are under the joint supervision of two supervisors. Effectively two supervisory styles are possible. The Likert [14] theories of supervisory style and the effect upon subordinate workers would suggest that this difference in supervisory style could lead to measurable differences in individual productivity drift.

*H*3. *Productivity drift is dependent upon location in the plant.* Certain locations within the same plant may be more attractive in respect of nearness to various amenities and physical comfort (noise, light, warmth, drafts, etc.) or in respect of psychological comfort (for example, proximity to particular colleagues or management staff). Although the whole division is located within one section of a large building, the sub-groups seemed to fall into 4 clusters and it is possible that these particular locations may have affected individual productivity drift.

*H*4. *Productivity drift is dependent upon age.* Calendar age is probably the most popular demographic variable used to explain productivity differences. The mean age of each productivity drift group was tested against the mean age of the other 2 groups for statistical independence.

*H*5. *Productivity drift is dependent upon the length of service with the company.* Longer serving members receive higher basic wages while bonus payment rates are common to all employees. Thus it is possible that since the bonus offered is less attractive (relative to basic wages), the incentive offered is less attractive to longer serving members and this could lead to differences in productivity drift. On the other hand, it is also possible that longer serving members tend to identify themselves more closely with the firm and feel more loyalty towards the firm which might also explain differences in productivity drift.

*H*6. *Productivity drift is dependent upon the frequency of daily job-changes.* This is in effect a job-enlargement type of factor. The variety in work may be welcomed as a change from monotony and thus lead to positive productivity drift.

*H*7. *Productivity drift is dependent upon attendance.* Absenteeism has been conclusively associated with job-dissatisfaction and a large number of studies have been conducted to seek a relationship between job-satisfaction and productivity [6]. It was thus thought desirable to test if there was any connection between attendance and individual productivity drift.

*H*8. *Productivity drift is dependent upon the individual claims for excess work.* From section 4.2 it is possible that productivity ratings are 'engineered' by the workers by skillfully using claims for excess work to determine individual productivity ratings.

6. RESULTS

On the basis of patterns of productivity drift, the 114 workers in the division may be divided into 4 groups based on drifts which are significant at least at the 0.05 level.

(i) Group 'a'—43(38%) of the workers exhibited "drift zero" over the whole of the study period of 82 weeks.

(ii) Group 'b'—14(12%) of the workers exhibited an initial "drift plus" followed by "drift zero" categorised irrespective of how many "drift pulses" are recorded.

(iii) Group 'c'—20(18%) of the workers exhibited "drift zero" followed by "drift plus", (in 17 cases "drift plus" was significant at 0.01 level, and 3 cases at 0.05 level) categorised irrespective of how many "drift zeros" are recorded.

(iv) Group 'd'—37(32%) of the workers exhibited at least one time interval (ΔT) of "drift minus" (in 21 cases "drift minus" was significant at 0.01 level and in 16 cases at 0.05 level).

The results of the hypothesis testing show the null hypothesis as having been rejected at the 0.05 level in three cases:

(i) Younger workers and also workers with fewer years of service in the job display significantly more instances of "drift plus".

(ii) Older workers and workers with more years of service in their current job display significantly more instances of "drift plus" and "drift zero".

(iii) Those displaying "drift zero" claim significantly more excess time than either the "drift minus" or the "drift plus" groups.

6.1. *Further tests on the drift pattern of group 'd'*

In 18 instances, negative productivity drift is preceded by zero productivity drift and in 4 instances by positive productivity drift; in 23 instances negative productivity drift is followed by zero productivity drift and in 10 instances by positive productivity drift.

To test if there is any relationship between the occurrence of negative productivity drift in any interval ΔT_n and the sign of the productivity drift during the interval ΔT_{n-1} preceding it or ΔT_{n+1} following it, the following null hypothesis was set up and tested using ψ^2 test. Null hypothesis: The occurrence of negative productivity drift is not significantly related to the ante- or post-occurrence of zero or positive productivity drift. Test at 0.05 level, degrees of freedom = 1, Reject: $\psi^2 \geqslant 3.85$. Calculated $\psi^2 = 1.79$. Therefore the null hypothesis is retained.

6.2. *Further analysis of excess claims in interval T_2*

In section 4.2 it was noted that claims for unmeasured work categorised as Excess Codes accounted for 80% of all claims for unmeasured work. Although Excess Codes include 32 specific claims, 72% of total unmeasured work in fact comes under 4 Excess Codes, namely,

(a) Operator retraining—27%
(b) Poor quality parts—17%
(c) Small orders and samples—16%
(d) Defective assembly—12%

The claims for unmeasured work are of considerable interest because the total amount of unmeasured work for the 3 drift groups in period T_2 was seen to be: Drift Plus—53% of total attendance; Drift Negative—45% of total attendance; Drift Zero—64% of total attendance. The results of further analysis of the Excess Codes show that the results for the "drift zero" group differ significantly at the 0.05 level.

7. DISCUSSION OF RESULTS

The pattern of productivity drift displayed by groups 'c' and 'd' can be represented on the productivity models shown in Fig. 2 only as finite discontinuities. Two types of discontinuities are possible:

(i) Discontinuity I—this pattern describes a temporary levelling off in productivity followed by a recovery to a new higher level of productivity. This pattern is similar to that already observed in the learning phase and which has been termed the 'plateaux effect' [8, 9, 20, 21]. This pattern has been exhibited by group 'c'.

(ii) Discontinuity II—this pattern, exhibited by group 'd', is apparently a sustained period of "drift minus" which can occur at any point in the drift phase of the individual productivity curves. There seem to be 2 patterns of recovery, one being a return to earlier productivity levels "drift zero" and

Vol. 6
1978

a second being a recovery to an increased productivity level "drift plus".

The hypothesis testing into possible causes for particular values of drift reveal an apparent difference between the group including younger workers and also workers who are comparatively new to the job and a second group including older workers and also workers who have had longer service in the same job. It has been shown in various studies for example, [4, 10] that calendar age is not a determinant of work performance. This observation together with the strong correlation between the two measures of time ($r = 0.48$, $P \leqslant 0.001$) suggests that the relationship with 'age' is spurious. Thus there seems to be a relationship between 'time in job' and drift—with workers who have been in the same job for longer periods showing a tendency to "drift minus" and "drift zero".

The second observation to be made from the results of the hypothesis testing is that workers who display "drift zero" claim significantly more excess time. Since "drift zero" is the long term steady state, it seems that workers who have been in the job for a longer period claim significantly more excess time. It is thus possible that this steady state is 'engineered' by experienced workers. This tentative conclusion agrees with the findings of Roy [16, 17] who carried out a detailed study of shopfloor work behaviour in manufacturing industry and described this pattern of work behaviour in terms such as 'systematic soldiering' and 'making out'—this is essentially cheating to maintain measured productivity at predetermined levels without actually achieving the corresponding outputs. Further testing of detailed claims under Excess Codes show that the most significant difference is in claims for Defective Assembly followed by claims for Poor Quality Parts and there is no difference in the claims for Operator Retraining which is the largest single claim at divisional level. This pattern of claims is interesting, because the two claims listed above are also the most difficult to countercheck thus supporting the hypothesis that these claims are being used by experienced workers to control their income. It follows from these results that conventional work measurement methods are ineffective on experienced workers.

8. CONCLUSIONS

The findings reported here show that two types of discontinuities are possible in the Extended Learning Curve. The first is similar to the plateaux effect already noted in the learning phase (Sriyananda and Towill [20]). The second is the occurrence of a sustained period of negative changes in productivity which can apparently occur at any point in the drift phase. It has also been suggested that the long term steady state of unchanging productivity patterns displayed by experienced workers is quite possibly the result of successful attempts at 'beating the system' and need not necessarily be indicative of true output.

REFERENCES

1. Bevis FW (1970) An exploratory study of industrial learning with special reference to work study standards. MSc Thesis, UWIST, Cardiff, UK.
2. Bevis FW, Finnear C & Towill DR (1970) Prediction of operator performance during learning of repetitive tasks. *Int. J. Prod. Res.* **8**(4), 293–305.
3. Bohlen GA & Barany JW (1976) A learning curve prediction model for operators performing industrial bench assembly operations. *Int. J. Prod. Res.* **14**(2), 295–303.
4. Bowers WH (1952) An appraisal of worker characteristics as related to age. *J. Applied Psychol.* **36**, 296–300.
5. Brayfield AH & Crockett WH (1955) *Employee Attitudes and Employee Performance, in Management Motivation* (Eds Vroom VH & Deci EL). Penguin, Harmondsworth, Middx, UK, pp. 72–82.
6. Camerson S (1973) Job satisfaction: the concept and its measurement. Work research Unit, Department of Employment, London.
7. De Jong JR (1957) The effects of increasing skill in cycle time and its consequences for time standards. *Ergonomics* **1**(1), 51.
8. Glover JH (1966) Manufacturing progress functions. I. An alternative model and its comparison with existing functions. *Int. J. Prod. Res.* **4**(4), 279–300.
9. Glover JH (1966) Manufacturing progress functions. II. A selection of trainees and control of their progress. *Int. J. Prod. Res.* **5**(1), 43–59.
10. Heron A & Chown S (1967) *Age and Function.* Little and Brown, Boston, Massachusetts.
11. Hindmarsh GW & Towill DR (1974) Theory and application of the time constant learning curve model. *Paper WA*16.14, ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico.
12. Hitchings B & Towill DR (1975) An error analysis of the time constant learning curve model. *Int. J. Prod. Res.* **13**(2), 105–135.
13. Kaloo U. (1977) Doctoral Thesis, to be presented to UWIST, Cardiff, UK.
14. Likert R (1961) *New Patterns of Management.* McGraw-Hill, New York.
15. Murrell KFH (1971) Industrial Work Rhythms, in *Biological Rhythms and Human Performance.* (Ed Colquohoun WP). Academic Press, London. Chapter 12.

16. Roy D (1952) Quota restriction and goldbricking in a machine shop. *Am. J. Sociol.* **57.**
17. Roy D (1953) Work satisfaction and social reward in quota achievement: an analysis of piecework incentive. *Am. Sociological Rev.* **18.**
18. Roy D (1955) Efficiency and 'the fix': informal intergroup relations in a piece work machine shop. *Am. J. Social.* **60,** 255–266.
19. Seymour WD (1966) *Industrial Skills.* Pitman, London, p. 89.
20. Sriyananda H. & Towill DR (1973) Prediction of human operator performance. *IEEE Trans. Reliability* **R22**(3), 148–156.
21. Towill DR & Bevis FW (1973) Managerial control systems based on learning curve models. *Int. J. Prod. Res.* **11**(3), 219–238.
22. Towill DR & Eler MS (1975) Performance models in a process industry. *Technical Note EP* 30, UWIST, Cardiff, UK.
23. Wright TP (1936) Factors affecting the cost of airplanes. *J. Aeronaut. Sci.* **3.**

ADDRESS FOR CORRESPONDENCE: *Professor DR Towill, UWIST, Department of Mechanical Engineering and Engineering Production, King Edward VII Avenue, Cardiff CF1 3NU, UK.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 305–311
© Pergamon Press Ltd 1978. Printed in Great Britain

0305-0483/78/0901-0305$02.00/0

Modeling and Solution of Practical Management Problems through Netforms

RANDY GLOVER

Management Science Software Systems, Boulder, USA

PAUL TALMEY

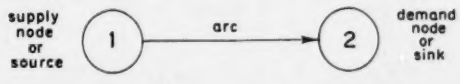
University of Colorado

(*Received December 1977*)

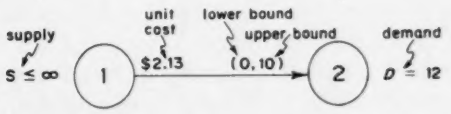
This paper shows some useful schemes to graphically depict netform models of practical inventory problems. We begin with a simple inventory problem and then progress to more comprehensive representations of an inventory system by network and netform model techniques. Some text books mention that very basic inventory problems may be modeled as networks. But they usually offer very little instruction in the actual development of a network inventory model, and invariably conclude that network models can't accommodate the complications of the vast majority of real-world inventory problems. Further, they completely fail to convey the modeling power available to more general netform (network-related formulation) techniques. As text books catch up to the recent innovations now being applied in practical settings, this deficiency will change. Our purpose is to introduce the fundamental ideas and to provide an understanding of the possibilities inherent in these model innovations. We begin from first principles, so that no prior familiarity with network model representations is necessary to follow the exposition. Because of a heavy reliance on pictorial illustrations (another of the unique advantages of networks and netforms), we have also avoided the use of complex mathematical notation, and develop the key concepts by informal discussion.

NETWORK MODEL NOTATION

THE BASIC elements of network representations are called arcs and nodes. The following represents a flow of units from a supply or source node along an arc to a demand or sink node.



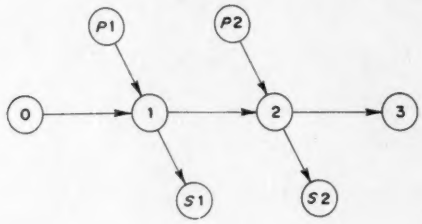
Flow of units along an arc usually has some cost associated with it and possibly capacity restrictions. The following network says there is an unlimited supply, a demand of 12 units, a 'shipment' cost of $2.13 per unit, and a shipping capacity of 10 units.



By convention the lower bound need never be negative. A negative bound (hence a negative flow) may be viewed as applying to a reverse flow. A reverse flow from node 2 to node 1 would be represented by another arc with its tail at node 2 and its arrow head at node 1. We could force at least two units to be sent from node 1 to node 2 by the bounds "(2,10)". As a rule, the bounds are not shown if the lower bound is 0 and the upper bound is infinite or redundant. (e.g., $(0,\infty)$ or, in the above case, (0,13).)

BASIC NETWORK INVENTORY MODEL

The following is a basic two period inventory maintenance model:



where:

Node 0 introduces any initial inventory level;

*P*1 is the source node for purchases in period 1;

Node 1 receives the total inventory in period 1;

*S*1 is the sink node for sales in period 1;

Arc (1,2) carries excess inventory into period 2;

*P*2 is the source node for purchases in period 2;

Node 2 receives the total inventory in period 2;

*S*2 is the sink node for sales in period 2;

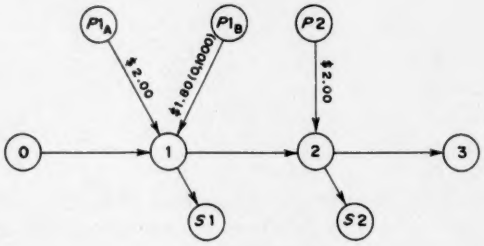
Node 3 receives any ending inventory level.

This is called a transshipment network since flow shipped to nodes 1 and 2 can in turn be shipped to other nodes. For transshipment nodes that have no net supply or demand requirements of their own, the amount of flow received at the node must equal the amount shipped out of the node. Thus in the diagram above, the initial inventory plus purchases in period 1 must equal the sales in period 1 plus the amount of inventory carried into period 2. Or as it is more often stated, ending inventory equals beginning inventory plus purchases less sales.

The objective is to determine how much to purchase and maintain in inventory in each time period to minimize total cost (or maximize total profit). Data for the problem would include a forecast of product demand (expected sales volume), expected purchase prices and inventory carrying charges. We will expand the model to accommodate real-world considerations such as multiple sources of supply, inventory capacity restrictions, maintenance of safety stock levels, back-order costs, budget limitations on purchases, quantity discounts on purchases (and/or fixed order or set-up costs), and multi-commodity inventories.

MULTIPLE SOURCES OF SUPPLY

This one is simple... we just add another supply source to our inventory node. The following models Source *A* offering as many units as we wish at $2.00 each, and Source *B* offering a maximum of 1,000 units in period one at a cost of $1.80 per unit.



INVENTORY CAPACITY RESTRICTIONS

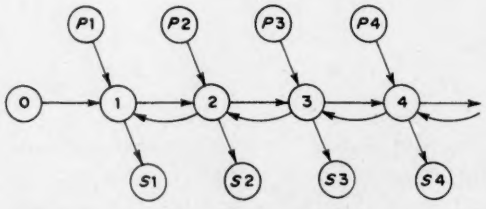
Another easy one. We simply place an upper bound on our carrying charge arc from one period to the next (arc (1,2) above). This one does not even deserve another diagram.

MAINTENANCE OF A SAFETY STOCK LEVEL

Now we place a lower bound on our carrying charge arc. This forces us to carry at least so many units from period to period, just in case....

BACK-ORDER COSTS

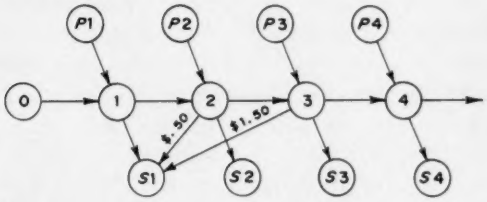
First the classic model:



Suppose *S*1 demand is greater than we can handle with purchases and existing inventory in period 1. The addition of the reverse flow

Vol. 6
1978

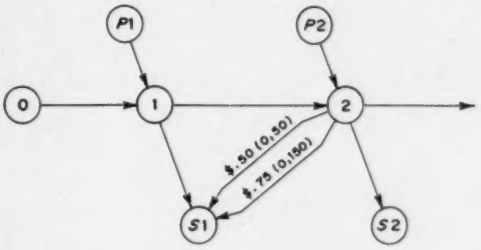
arcs permit us to use as much of period 2 inventory as we need to serve period 1 demand. At first glance, this seems O.K. After all, we can introduce whatever cost we want on the reverse flow arcs to represent both the actual additional handling costs and some value for customer dissatisfaction. But what if we want to limit back-orders to one or two periods at most? With the model above, we could arrive at a solution filling period 1 sales out of period 10 inventory. What if it costs more per unit to handle 100 back-order items than it does to handle only 50, or vice versa? What if we know that historically 20% of our customers cancel their orders when faced with a back-order? We can, and will, model all of these situations.



Arcs (2,S1) and (3,S1) represent shipping to period 1 sales from period 2 and period 3 inventory. We have limited back-orders, in this case, to span at most two periods. Another arc has been added, but we won't find ourselves filling period 1 sales out of period 10 inventory. And we have the added capability of being able to place different costs on back-orders as they age, possibly representing increasing customer dissatisfaction or actual cost differences. (Another alternative is to draw the back-order arc directly from P2 to S1. This may even be preferable in special cases such as filling back-orders by having units drop-shipped from the supplier directly to the customer, or where there are special cost considerations wholly unrelated to normal order costs.)

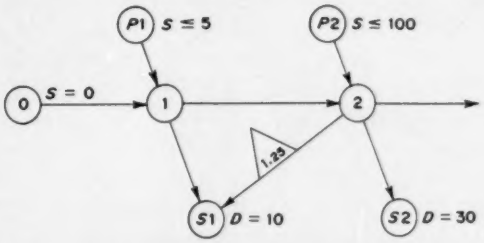
Notice how easily and simply these modifications were observed and effected with a network formulation. With only a mathematical formulation to go by, the modeler would likely be well into program testing before finding such simple improvements. It is particularly worth noting that standard text-book inventory models do not allow the simple (but crucial) limitations on back-orders that are handled in a natural and elementary fashion in this example.

Now let us expand our model.



Here we have first period back-order costs of $.50 per unit for the first 50 units back-ordered and $.75 per unit for back-orders in excess of 50 units, with an overall limit of 200 units permitted to be back-ordered. Because the objective is to minimize costs, the program will always fill up the less expensive arc before allowing any flow on the more expensive arc. If we wanted to charge more for the first 50 units on back-order, we could do this also, using the techniques described later for price discounts and fixed order costs.

How do we handle order cancellations? Is it possible to order only 9 units to satisfy a demand of ten units without upsetting the mathematical balance of the equations? To do this, we leave networks and step into the realm of generalized networks, the first step toward the power of netforms, with the use of an arc multiplier. Consider the following:



In this example, we have no initial inventory, purchases in period one are limited to at most 5 units, and we expect 20% of back-orders to be cancelled. We run the model with a sales forecast of 10 units in period 1. We can see that our limit on purchases in period 1 will force us to place 5 units on back-order. But if we do that, we will get a cancellation of 20% (1 unit). The practical result is a running change in demand while the program is executing. By placing a multiplier on arc (2,S1) we can, in effect, ship 4 units from period 2 inventory and satisfy the demand for 5 units of

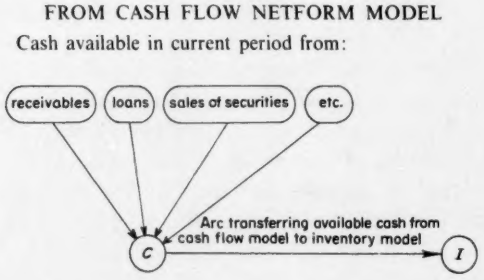
period 1 sales. The value of our multiplier in this example would be $1/(1-.2) = 1/.8 = 1.25$. The cost on (2,S1) must be high enough that the loss of sales associated with flow on this arc is avoided like the plague (whatever value we decide to place on the plague).

To cover the possibility that demand in some period may be greater than we can supply, we simply insert a dummy source node with unlimited supply and an arc to each demand node. Each arc would have a penalty cost great enough that it would have flow only after all other sources of supply were exhausted.

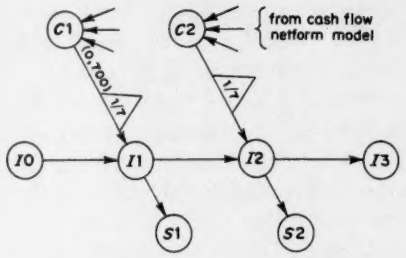
BUDGET LIMITATIONS ON PURCHASES

Our financial planning V.P. has just told us that we only have \$700 to spend on inventory purchases this period. Assuming we know the unit cost of inventory, we can translate this into an upper bound on the total number of units purchased. For example, if inventory costs \$7 per unit, then the \$700 restriction translates into an upper bound of 100 units. This bound may then be attached to the arc (*P*1,1) in the preceding diagram.

However, in a more complex situation, the limit on available cash to purchase inventory may not be a predetermined constant, but rather a variable determined by integrating the inventory netform model with an associated cash flow netform. To illustrate, suppose we have the situation depicted in the following diagram:



Note that flow entering node *C* is in units of cash (i.e. dollars), while the flow entering node *I* should be in units of inventory. At \$7 per unit, a multiplier of 1/7 will 'transfer' our purchase dollars into units of inventory.



It would alternatively have been possible in the preceding example to put multipliers on each of the several arcs from the cash flow netform leading to the inventory netform, doing away with the intervening "cash to inventory transfer arc" denoted by (*C*, *I*). On the other hand, the intervening arc makes it possible to put a collective limit on cash from all sources that may be spent on inventory. Note that if different cash sources were to be allocated to different inventory purchase options, then this could be handled by the use of multiple transfer arcs, or simply by directing appropriate subcollections of arcs from cash sources to respective inventory options. Finally, the sales from the inventory netform generate cash and receivables that provide inputs back into the cash flow netform, using multipliers to convert units of inventory back into cash. (In some situations, it may be as convenient to express the entire inventory netform model in terms of cash flows rather than inventory flows. However, multipliers would still be required to handle the fact that unit purchases costs are different from unit sales costs.)

This example of an integrated cash flow and inventory model is of course highly simplified for ease of exposition, but the fundamental constructions may be used to model a variety of more complex interactions as well as to provide constructions for the cash flow netform itself. Rather than elaborate these, we will continue to focus on the inventory model, and show how to handle additional important elements of inventory management by netforms.

Further extensions of generalized network models into the capabilities of netforms will include one or more of the following netform features:

1. Placement of integer restrictions on particular arcs.
2. Application of a 'fixed charge' to an arc in addition to some variable or per-unit charge.

Vol. 6
1978

3. Designation of an arc as an 'all-or-nothing' arc (flow on the arc must be either zero or at its upper bound).

4. Introduction of multiple choice conditions. With netforms, we can specify two arcs as "either/or" arcs where either one arc or the other arc may have flow, but not both. We can expand this with more than two arcs to say "at most one arc may have flow" and even "at least one arc must have flow."

5. Specification of 'side constraints' such as those which express proportional relationships between flows.

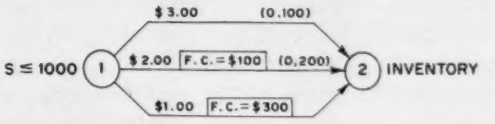
In our next examples we illustrate two very important aspects of netforms. First, we show how they can be used to model price discounts on purchases (and, by extension, for other non-linearities involving increasing returns to scale, or combinations of increasing and decreasing returns to scale). Second, we provide alternative model constructions to handle the same basic situation, demonstrating the ability to organize different netform components to achieve the same objective. This provides greater flexibility for the practitioner to design his models to take advantage of solution algorithms specifically tailored to particular types of constructions. It also gives him flexibility to design his models in a way that best reflects his feel for the problem, and to obtain insights into alternative ways of viewing its underlying components.

PRICE DISCOUNTS

The next four diagrams will be of the supply portion of our inventory model (arc (P1,1)) expanded to accommodate quantity discounts on purchase price. Three of the models require the placement of an upper bound on the quantity of units available. If there is no restriction on supply availability, any non-limiting upper bound may be selected. Data for all models is:

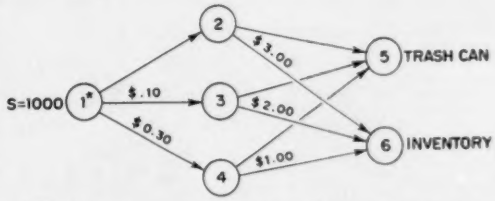
0–100 units @ $3.00 per unit;
101–200 units @ $2.00 per unit;
201 & up units @ $1.00 per unit;
1000 units maximum in any period.

Our first model uses the 'fixed charge' netform capability.



The value of the fixed charges for the middle and lower arcs are the calculated differences between the discounted unit cost sum and actual cost at the number of units where the discounted unit cost begins. There need be no bounds on the $3.00 and $2.00 arcs above because the least cost solution technique will automatically direct flow from one arc to another as unit flow increases and subsequent arcs become less expensive. The addition of bounds to these arcs, however, makes the model more readable.

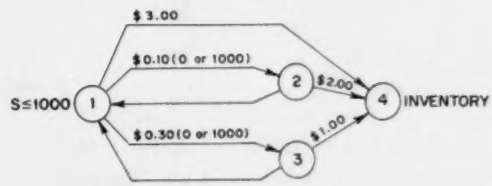
The following diagram effectively models the same situation with the introduction of the "=" symbol on the supply node and the specification that only one arc leaving the supply node may have flow.



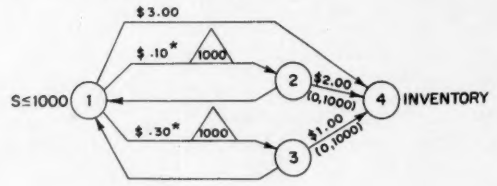
* At most one arc leaving this node may have flow.

Because the "=" symbol forces 1000 units to leave the supply node, a dummy or trash-can node is introduced to accept units leaving supply but not taken into inventory. In a zero purchase situation, the no-cost solution would be to send all 1000 units along arcs (1,2) and (2,5) to the trash can.

A similar approach is to permit flow to be zero or at its upper bound and then allow any surplus to be returned. The following is such an "all-or-nothing" model. The representation of bounds as "0 or 1000" means exactly that: flow may be either 0 or 1000.

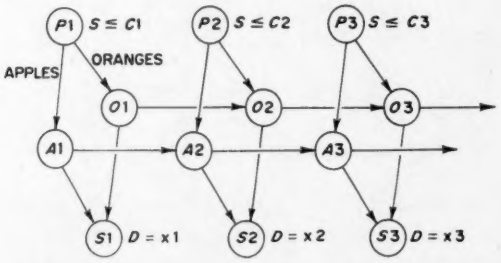


The next model accomplishes the same result using the integer restriction and multiplier features. It is important to notice the order in which symbols are encountered along an arc. The integer restriction symbol (*) is encountered before the multiplier, and the program would use the same sequence in applying tests. Placing the integer restriction symbol after the multiplier is permissible, but would have a quite different effect.



MULTI-COMMODITY MODELS

A common multi-commodity restriction is a limitation on the total amount of purchases or of shipments of the different commodities. The following is a two-commodity inventory (apples and oranges) with limits on both purchases and shipments.



Another multi-commodity restriction might be an overall storage capacity limit on the different commodities. In the above model, we would add the restriction that the flows $(A1,A2) + (O1,O2) \leq K$.

Any combination of netform modeling techniques presented in the single-commodity framework may also be applied to the formulation of multi-commodity models.

ADVANTAGES OF NETFORM MODELING

The underlying network and network-related structures of netform models can be exploited by special algorithmic procedures that are vastly more efficient than those designed for solving the same problems when cast in general mathematical programming formulations. Real-world problems may contain thousands of variables and constraints, making streamlined solution techniques of the type available to network formulations a practical necessity.

The simplicity of the graphical representation of netforms enables a clearer understanding of the interrelationships of activities by both manager and modeler alike (as noted in the previous section). This promotes a more precise formulation which often results in superior solution quality as well as a faster solution. And this readability of the netform model by management should bring the OR man back into the management community with resultant increased cooperation and practical benefits.

Classical inventory models, which we have been obliged to live with for years, generally amount to a set of complicated decision rules to tell how much to order (or build) and when to order it, based on certain assumptions. There are different inventory models for different sets of assumptions. And the models themselves apply *local* decision rules at various steps in the problem, which, taken together, result in a final "solution."

A netform model of the inventory problem offers, in addition to benefits listed above, a *global* decision rule which presents a least-cost solution independent of assumptions. In classical inventory models, a change of assumptions requires a different model. Moreover, many assumptions of classical models are *essential* to the model (though unfortunately, often unsatisfied by conditions in the real world) in order to enable the classical ordering formulas to apply. Netform inventory models do not require many of these limiting assumptions. Thus, for example, they are not dependent upon assumptions concerning distribution of demand (e.g. whether it is monotone increasing or decreasing), or on different supply or storage expectations. A revised forecast is simply input just as the original data was.

Individuals not familiar with classical inventory models find it intuitively much easier to accept the concept of a single model framework that embraces a multiplicity of supply, demand and storage interrelationships than to confront a battery of different formulas, each (partly) applicable to a separate set of condi-

Vol. 6
1978

tions, or to different items within the product line. This ready acceptance of the network inventory model by the uneducated will surely result in greater managerial acceptance and implementation. To solve the model, the manager needs only to supply a demand forecast and source availability information. The model will then solve for the optimal solution regardless of rising, falling, or static demand, price discounts, ordering or set-up costs, capacity limitations, or back-order allowances. Simple local decision rules from classical models may seem easy to apply, but if there is a difference of even a few percent between local and global netform optimality, there may be large cash savings available from a global netform model that are wasted by reliance on out-of-date techniques. Further, purchase order instructions that are generated from the result of a computer netform optimization need not be more difficult to apply than those generated from other sources. Indeed, the interactive capabilities being incorporated into the highly efficient netform solution programs enable decision makers to pose useful "what if" questions that additionally enhance the value of netform models. The clarity of these models, and their lack of reliance on specific assumptions, invites rather than stymies the posing of such questions.

SUMMARY (OR, WE SAY AGAIN...)

The netform inventory problem may be modeled for periods of any length desired, or for periods of unequal lengths. No averaging or smoothing of demand is necessary. The model permits different order costs, unit costs, carrying costs and back-order costs at each period if desired. Back-order costs may be increased with the age of the back-order. Budget limitations on purchasing may be introduced on any period, or on several periods in differing amounts. Capacity limitations on production or storage facilities are easily accommodated. A safety stock level is easily imposed. Multiple sources of supply and quantity discounts fit right in. Text book inventory models allow for very few of these features, even taken in isolation. Netform models allow for all of them, taken in *conjunction*. Finally, their structural form and speed of solution enables the use of dynamic sensitivity analyses and the posing of "what if" questions, which can frequently be handled in a real-time, interactive, man-machine environment.

Consideration should be given to the possibility that network inventory models just might be that good!

REFERENCES[1]

1. BHAUMIK G (1973) *Optimum operating policies of a water distribution system with losses*. Unpublished Dissertation, The University of Texas at Austin.
2. CHARNES A, GLOVER F, KARNEY D, KLINGMAN D & STUTZ J (1975) Past, present and future development, computational efficiency, and practical use of large-scale transportation and transshipment computer codes. *Computers and Op. Res.* **2**(2), 71–81.
3. COLLINS M, COOPER L & KENNINGTON J (1976) Solving the pipe network analysis problem using optimization techniques. *Technical Report IEOR 76008*, Southern Methodist University, Dallas, Texas.
4. CRUM R (1976) *Cash management in the multinational firm: a constrained generalized network approach*. Working Paper, The University of Florida, Gainesville.
5. EVANS J (1975) *A single commodity network model for a class of multicommodity dynamic planning problems*. Working Paper, University of Cincinnati.
6. GILLIAM G & TURNER J (1974) *A profile analysis network model to reduce the size of microdata files*. Working Paper, Office of Tax Analysis, Office of the Secretary of the Treasury, Washington, D.C.
7. GLOVER F, HULTZ J & KLINGMAN D (1977) Improved computer-based planning techniques. *MSIS 77-5*, School of Business, University of Colorado, Boulder. (to be published in *Interfaces*).
8. GLOVER F & KLINGMAN D (1975) Real-world applications of network related problems and breakthroughs in solving them efficiently. *ACM Trans. on Mathematical Software* **1**(1), 47–55.
9. GLOVER F, KLINGMAN D, MCMILLAN C and TALMEY P (1977) Netforms: new advances in mathematical modeling. *Proc. of First Int. Conf. in Mathematical Modeling*. University of Missouri at Rolla, Missouri.
10. KAMERSKY P (1974) A computer code for refueling and energy scheduling containing an evaluation of nuclear decisions for operation. Unpublished PhD Thesis, Ohio State University.
11. KIM Y (1972) An optimal computational approach to the analysis of a generalized network of copper refining process. Presented at the Joint ORSA/TIMS/AIIE Conference, Atlantic City, New Jersey.
12. SMITH MCGREGOR J and LIEBMAN JS (1977) A zero-one integer programming formulation of the land use assignment and transportation network problem. *ORL/77/7*, University of Illinois at Urbana—Champaign.

[1] Many of the references containing good (and up-to-date) network-related modeling ideas are unpublished, with a few noteworthy exceptions. This list of references indicates some useful sources for the reader interested in such modeling ideas in other real-world applications.

ADDRESS FOR CORRESPONDENCE: *Randy Glover Esq, Management Science Software Systems, 1040 Lehigh Street, Boulder, CO 80303, USA.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 313–323
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0313$02.00/0

Mixed-Model Assembly Line Sequencing Problems

EZEY M DAR-EL

Israel Institute of Technology

(*Received December 1977; in revised form February 1978*)

The paper develops a comprehensive classification of mixed-model assembly lines from which four categories of model sequencing are derived. Each category aims at satisfying one or both of two objective criteria, the one minimizing the overall line length, and the other minimizing the thruput time. Approaches for solving the sequencing problem in each category are presented. The paper suggests a design strategy that can be followed by designers of mixed-model assembly lines. Specific topics requiring further research are also defined.

1. INTRODUCTION

THE MIXED-MODEL line sequencing problem was first enunciated in 1963 by Kilbridge and Wester [4], yet, until today, some 14 years later, only a few papers on this topic have appeared in professional journals. While mixed-model lines are found in many manufacturing plants, a practical method for tackling the attendant sequencing problem has still to be developed for adoption by industry. However, a comprehensive classification of mixed-model assembly lines is needed in order to clearly distinguish between the different types of sequencing problems that may arise.

This paper presents a classification of mixed model assembly lines and their main design features. Various approaches for solving the appropriate sequencing problem are discussed.

2. CHARACTERISTICS OF MIXED-MODEL LINES

Many types of Mixed-Model assembly lines exist in industry [7] and developing a classification system is best approached through identifying and distinguishing between several important design features. These are discussed below.

(1) *The conveyance system*

This can either be a *conveyor* moving at uniform speed, or else a *stationary system* where the product remains stationary at each station, and only transferred to the next station after completion of the station's task. Alternatively, the next station's operators may replace the current ones, as in the assembly of large products, such as with aircraft, ships, etc.

(2) *The product's link to the conveying system*

Specifically, the question is whether or not the product can be moved independent of the conveyor movement, i.e. if the product can be removed from the conveyor, or else held stationary relative to the conveyor movement. Products that cannot be moved independently of the conveyor movement are referred to as being *Products Fixed*, whereas those with independent movement are called *Products Movable*. Products that are fixed may be rotated in any direction but may not be removed from their respective positions on the conveyor. Large products, such as washers and refrigerators, would be included in this category since these are often simply placed on the conveyor and rely on their weight to maintain their respective locations. Usually, there is no provision for buffer stock between stations in 'Products Fixed' assembly lines where operators

generally work under *paced* conditions. On the other hand, 'Products Movable' lines allow buffer stock to accumulate between stations, with the consequence that operators work under *unpaced* conditions.

(3) *The stations*

Stations can be *closed* or *open*. It is undesirable, or else impossible, for operators from adjacent stations to violate the boundaries of a closed station (e.g. the station could be a pit area, a spray paint booth, a dip tank, heating chamber, and so on). Therefore, the times available for completing work tasks allocated to stations in a 'Product Fixed–closed station' line, are constrained. Open station boundaries can be crossed, but the extent of the crossing can be either *restricted* (e.g. due to the limited range of powered tools, materials handling equipment, etc.) or *unrestricted*. In both cases, the requirement is that no interference occurs between adjacent operators. Clearly, there is flexibility in the times available for completing tasks allocated to open stations. A stationary line is an example of an assembly line having open stations, since a product 'completed' at one station can immediately be collected for processing by the downstream operator.

(4) *The launching discipline*

Kilbridge and Wester [4] described two launching disciplines: The first is *fixed rate launching* (FRL), where the launching period γ is the weighted average of the total assembly time for all products to be assembled, over all stations. The second discipline is *variable rate launching* (VRL), where the launching period is the first station's task time of the last product launched, i.e. on completing work on the current product, the first station's operator can immediately begin working on the next product. Selecting between these two launching disciplines, implies that *no* buffer stock exists between work stations. The launching discipline becomes irrelevant to the analysis when buffer storage is allowed. All researches on this topic have adopted the FRL (fixed) discipline for their models, assuming it to be the more practical of the two [1, 2, 5]. Not only is this unproven, but the VRL (Variable) discipline is used quite frequently with manual lines [7].

3. THE MIXED-MODEL SEQUENCING PROBLEM

Two approaches have been proposed so far for solving the mixed-model sequencing problem. One is by Thomopoulos [5, 6] who assumes a given facility (station lengths) and finds the sequence that minimizes the total penalty for a number of "inefficiencies" (operator idle time, utility work, work congestion and deficiency). This approach fails on two counts: it provides no help for determining station lengths (i.e. for a new design), and depending on the assumed dimensions of the station lengths, *any* sequence can be made to look good!

The second approach was proposed by Dar-El and Cother [1]. They select the sequence that minimizes the overall assembly line length for no operator idle time (discussed later) or other inefficiencies. This approach also determines the major facility dimensions (i.e. station lengths). Their method transforms all task processing times into operator movements on the assembly line, so that station lengths can be determined. However, neither approach makes the distinction between two objective criteria which can be identified in mixed-model sequencing problems.

Objective I to find the sequence that minimizes the overall facility length (cf. [1]); and

Objective II to find the sequence that minimized the thruput time (i.e. the typical flow-shop sequencing problem).

The physical description of these two objectives is illustrated in Fig. 1 which shows the operator movement for a closed station assembly line using the FRL discipline. If the conveyor speed is taken as unity, then the task processing times can represent the operator movement, i.e. the movement of the product fixed to the conveyor. Station lengths are then found by placing boundaries on the extreme operator movement locations (e.g. as with station 1 in Fig. 1). Actual station lengths are found as follows:

if

V_c = the conveyor speed (ft/min);

Vol. 6
1978

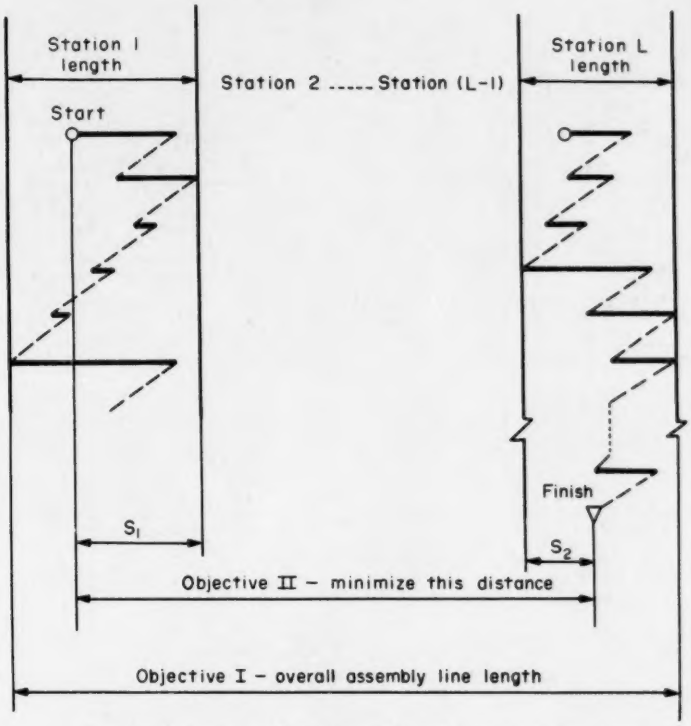


FIG. 1. *Objectives in solving the mixed-model sequencing problems.*

V_0 = the walking speed of the operator (ft/min), i.e. walking upstream to start work on the next product;

L = the number of work stations;

γ = the launching period (min.);

SLONG = the overall line length (min.),

then SLONG* = the actual line length, given by

$$\text{SLONG}^* = (\text{SLONG})\, V_c + \frac{\gamma V_c^2}{(V_0 + V_c)}. \qquad (1)$$

The second term in the equation accounts for the time taken for operators to walk upstream to start work on the next product.

Minimizing the overall facility length does not necessarily minimize the thruput time. Consequently, a trade-off between these two objectives may need to be considered when determining the minimum cost solution for a particular problem. Figure 2 illustrates the independence of those two objectives. Data for the sequencing problem given in Table 1 includes 3 models involving a total of 7 products to be processed over 2 stations.

$$\text{FRL period } \gamma = \frac{42}{7 \times 2} = 3$$

Figures 2(a) and 2(b) respectively represent the optimal sequence for Objective I (A.A.B.B.B.B.C) and Objective II (B.A.B.C.B.A.B), and are drawn for the open station case with products fixed to the conveyor (see Section 4). It is observed that overall line lengths and thruput times have different values in the two configurations.

4. A CLASSIFICATION OF MIXED MODEL PROBLEMS

A proposed classification of mixed-model sequencing problems is shown in Fig. 3. It ranges from conveyor lines on which products are fixed with closed stations and no buffer capacity, to lines on which the products are movable, buffer stocks exist (VRL discipline) and station boundaries can be crossed by adjacent operators.

It is observed that when products are not fixed to the conveyance system, there is no need to be concerned with station lengths.

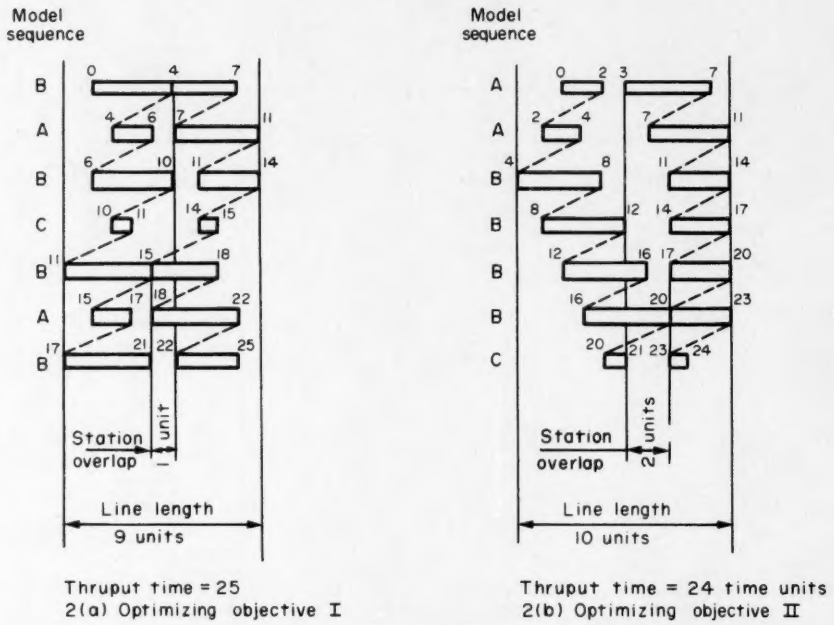


FIG. 2. *Optimal solutions for Objectives I and II–'products fixed–open stations' conditions.*

Whenever a product is completed at one station, it immediately becomes available for the downstream operator, who either straightaway works on it, or else, the product is held in temporary storage. Under these conditions, the station lengths should be determined according to proper workplace design principles and the mixed-model problem is reduced to one of finding a sequence that minimizes the thruput, i.e. the flow-shop sequencing problem.

On the other hand, both objectives apply to problems included in the 'Fixed' product category and cost trade-off conditions between these two objectives may need to be investigated.

Figure 3 also suggests that the station types are either all open, or all closed. This is not the case in practice, since station mixes (i.e. hybrids) do occur. The case of hybrid lines is discussed later.

5. ANALYSIS OF MIXED-MODEL SEQUENCING PROBLEMS (see Fig. 3)

(a) *Products fixed to conveyor–closed stations–VRL and FRL disciplines*

Both objectives defined earlier apply to this category, but it will be argued that minimizing the facility length is an overiding consideration, thus eliminating the need to consider the influence of thruput times.

The effect of closed stations is to make it *very unlikely* for the completion time of a product on one station, to coincide with the start time for work (on the same product) at the next station. For example, Fig. 4(a) illustrates the closed station version of the case given in Fig. 2(b). The overall line length is

TABLE 1. DATA FOR A SIMPLE MIXED-MODEL SEQUENCING PROBLEM

Model	Products required	Task times allocated to: Station 1	Station 2	Total assembly time
A	2	2	4	12
B	4	4	3	28
C	1	1	1	2
Total	7			42

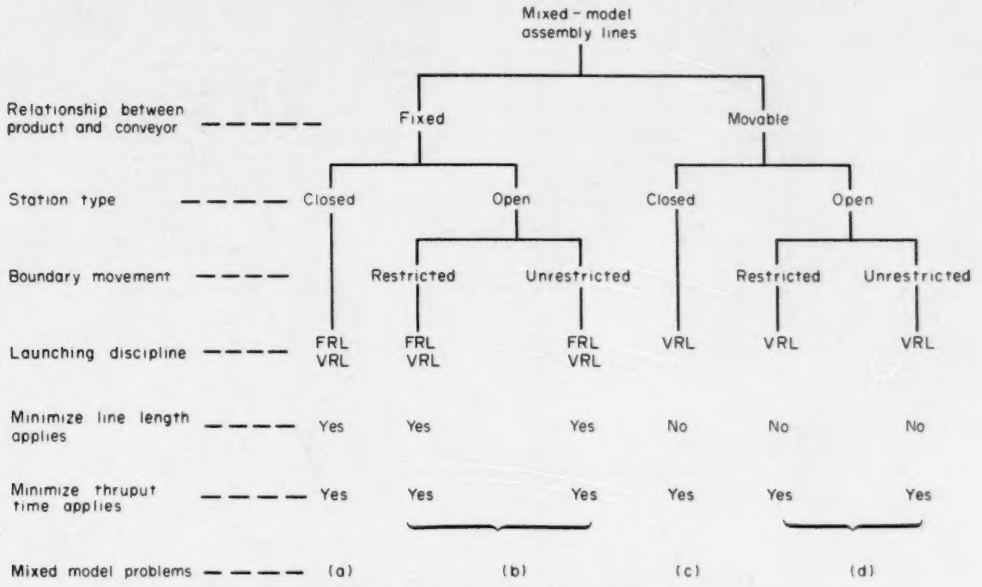


FIG. 3. *A proposed classification of mixed-model sequencing problems.*

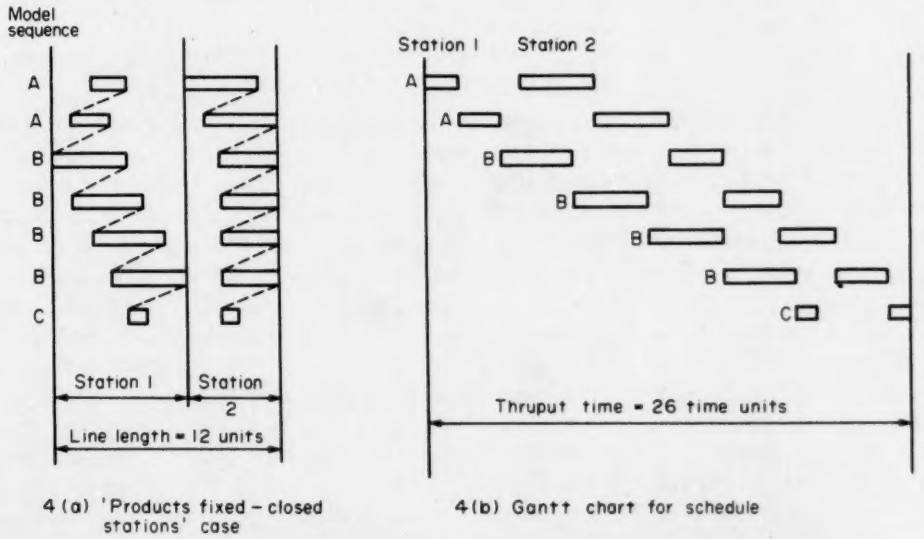
increased to 12 units (from 10 units) and the thruput time is increased to 26 time units (from 24 time units). Figure 4(b) shows the Gantt chart version Fig. 4(a) and illustrates the point, that products are unlikely to be worked upon continuously by adjacent operators.

Figure 4(b) also illustrates another feature of mixed-model lines with closed stations. For any station, once work is started, that station remains *continuously* occupied until all products are completed, i.e. with no operator idle time. The price paid for this 'ideal situation' is for station lengths to be increased until continuous operations can be maintained for each station.

Where there are many stations in the assembly lines, it becomes evident that a reduction in thruput time is better effected by station compression, i.e. by satisfying objective I only. Thus, referring to Fig. 1, to minimize TP (the thruput) means:

$$\text{Min(TP)} = \text{Min}\left\{S_1 + \sum_{i=2}^{L-1} (\text{Station length})_i + S_2\right\} \qquad (2)$$

where L is the maximum number of stations.



4(a) 'Products fixed – closed stations' case

4(b) Gantt chart for schedule

FIG. 4. *Illustrating the 'closed station' version of Fig. 2(b).*

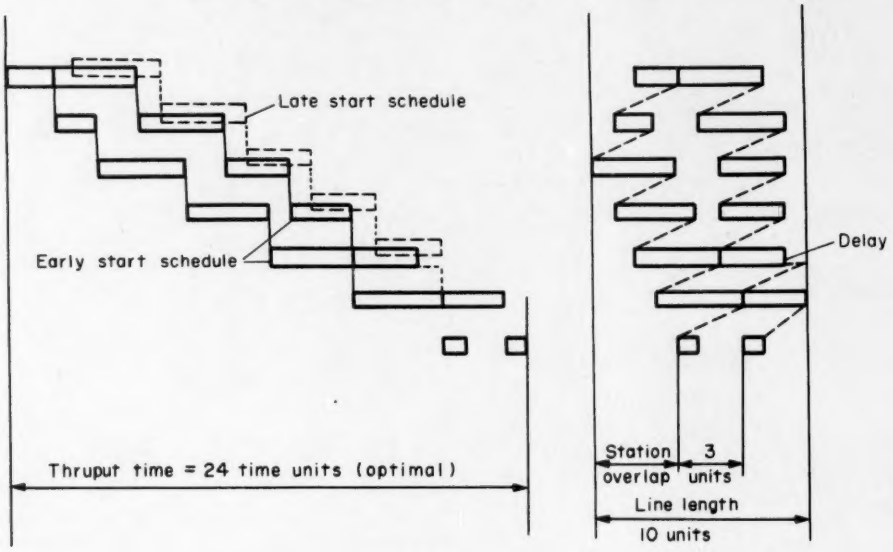


FIG. 5. *Open station version of Fig.* 4 (*see also Fig.* 2(*b*)).

Equation 2 is only marginally different from the expression for minimizing SLONG, the overall line length (Objective I), which is given by:

$$\text{Min(SLONG)} = \text{Min}\left\{\sum_{i=1}^{L} (\text{Station length})_i\right\}. \quad (3)$$

Consequently, it is assumed that the sequencing solution to 'Products Fixed—Closed Stations' problems should only aim at minimizing the overall station length. If L were small (say 4 or 5 stations), then a trade-off between the two objectives may need to be considered, as discussed, under problem (b) below.

The Dar-El–Cother model [1] is suggested as an appropriate method for solving this problem (a brief description is given in the Appendix). Fairly good results were obtained in applying this sequencing model to large lines—the final overall line lengths measuring between 5 and 15% above the appropriate $(\text{SLONG})_{LB}$ dimension. The Dar-El–Cother model can be used with either the FRL or VRL disciplines, there being no evidence to suggest that any one method is superior to the other. However, the FRL discipline is more likely to be used since it better suits the characteristics of several conveyor types (e.g. overhead and tow-line conveyors). It is an appropriate method to use when several lines are synchronized to feed into a main line which is used for the assembly of some large product (e.g. motor cars). On the other hand, if all models are approximately well balanced over all L stations, then there are advantages in choosing the VRL discipline, since *any* sequence would minimize both the overall line length as well as the thruput time. However, in the final analysis, the choice between the two disciplines may well be determined by the *cost* of the conveyor system.

(b) *Products fixed to conveyor–open stations–VRL and FRL disciplines*

For this category, both Objectives I and II should be considered for determining the assembly sequence, since each operation can begin at its *early start* time. On a Gantt chart this means that all jobs are moved as far to the left as possible as in Fig. 5(a) which illustrates the open station solution for the problem in Fig. 4(b). The solution shown in Figs 5(a) and 5(b) are identical to that in Fig. 2(b), although operations in the latter are drawn at their *late start* positions (as per the dotted schedule in Fig. 5(a). However, the Fig 2(b) version is preferred since it simplifies the analysis for the 'restricted movement' case (discussed later in this section).

The Dar-El–Cother algorithm for open stations by itself is inadequate, since no account is taken of the thruput time. Instead, the following approach is proposed.

First, determine a relationship equating the incremental cost of extending the facility length by one unit, with its equivalent value in reduced thruput time (i.e. increased production).

Let

P = total production required over a planned period;

Δt = time units (e.g. in centi-mins) be the equivalent value in thruput time for one unit of the facility length;

u = the unit labor cost (\$/hr);

F = the incremental cost per unit (0.01 V_c if time is measured in centi-mins) of facility length; F includes the additional facility cost and floor space required, as well as the extra power need to drive the system;

$u.P.\ \Delta t/6000$ is the \$ savings when TP is reduced by Δt (centi-mins). Then $\Delta t = 6000F/u.P.$

Example:

If

P = 10,000 products, produced over 1 year.

u = \$10 per h.

V_c = 4 ft. per min.

Cost per ft. of facility plus floor space taken, is estimated at \$2,500.

Then $F = 0.01 \times 4 \times 2500 = \$100$ per 'unit' of facility length. Hence, $\Delta t = 6$ time units (i.e. 0.06 min). Thus the cost of reducing SLONG by 1 unit is equivalent to reducing the thruput time by 6 units

The second step is to generate a number of solutions, each one having an overall line length (SLONG) and its associated thruput time (TP). These are plotted on a 'SLONG–TP' diagram (as in Fig. 6) and the first point that touches an upward moving 'equivalent cost' line provides the minimum cost solution.

Several methods can be used for generating solution sequences—some directed towards meeting Objective I while others, to Objective II. The Dar-El–Cother algorithm for open stations can be used as an example of the first method. Its results may be improved if the 'selection heuristic' can somehow take into account lower bound overall time estimates, such as used in solving flow-shop problems.

Sequences which satisfy Objective II could be obtained through any of the methods used in solving flow-shop problems, though the size of most industrial mixed-model problems may necessitate the use of heuristic methods. Treating mixed-model sequencing as a flow-shop problem, implies that the VRL discipline is utilised. The FRL discipline simply introduces unnecessary constraints [4] which requires that for any product in the sequence at any station i:

$\gamma \leqslant$ average task time for all launched products up to (and including) station i.

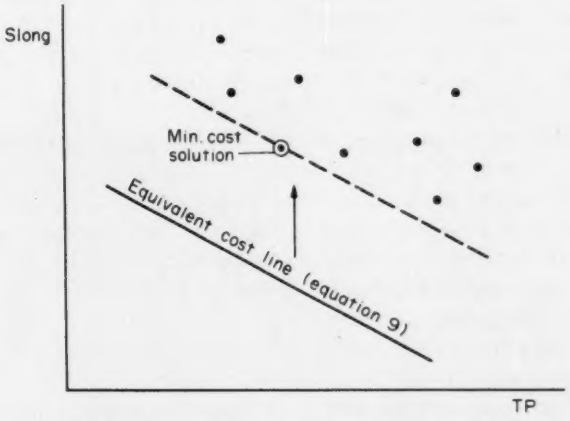


FIG. 6. *Relationship between SLONG and TP.*

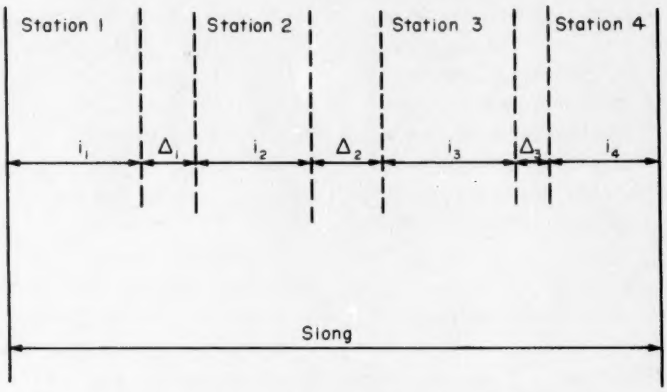


FIG. 7. *Illustrating the 'products fixed–open stations' case.*

The restricted movement case can be readily obtained from the unrestricted movement solution. Simply ensure that the overlap dimension (Δ_i) between adjacent stations (i and $(i + 1)$), is less than or equal to the allowable movement by elongating the line length where necessary. This procedure is illustrated in Fig. 7 which is drawn for a 4 station line.

If Δ_i overlap distance between stations i and $(i + 1)$ and design requirements specifies that the allowable movements be d_i, $d_i > 0$, $i = 1,2\ldots, (L - 1)$.
Then the 'corrected' overall line length, SLONG', would be:

$$\text{SLONG}' = \text{SLONG} + \sum_{i=1}^{L-1} \Delta_i - d_i$$

(only $\Delta_i > d_i$ terms included) (5)

(c) *Products movable–closed stations–VRL discipline*

The overall line length is not influenced by the sequence in cases where the product is free to move independently of the conveyor. This is because the product can always be held stationary upstream of an occupied work station. Station layouts are therefore designed according to proper plant layout principles.

This category is appropriate to the assembly of smaller and less expensive products which can be held as buffer stock between stations. It also applies to larger products when the conveyor system allows these to be held stationary at each station, e.g. with some types of tow-line conveyors. Closed station conditions (pit area, spray chamber, etc.) may also apply. Best results would be obtained with the VRL discipline. In fact, the FRL discipline would straightaway be 'transformed' to the VRL discipline since the products can move independent of their launched positions on the conveyor.

Assembly sequences for problems in this category are determined according to Objective II which corresponds to flow-shop problems. Thus, any algorithm that can produce good solutions to large flow-shop problems can also be utilized for mixed-model sequencing. One possible outcome is that the solution may require the segregation of all models, i.e. all products of one model assembled first, followed by all products of a second model, and so on. (This is likely to occur when each model is evenly balanced over the L stations.) There could be some advantage to production if this arrangement could be tolerated from marketing and warehousing considerations. This is because the 'learning factor' could play a dominant role in reducing performance time, when all product A's are assembled consecutively as against their being distributed over the production period (a day, 2 days, or, even a week).

The only effect on the analysis that closed stations has, is to increase the thruput times, i.e. to include the time taken for a 'completed' product to arrive at the downstream boundary. Closed stations have no influence on the assembly sequence.

(d) *Products movable–open stations–VRL discipline*

This arrangement gives the maximum flexibility in the design for mixed-model assembly.

Vol. 6
1978

Operators can cross upstream stations boundaries to collect and transfer the next product for processing at their own stations. This category also applies to the 'stationary' assembly line.

Stations are again designed according to proper plant layout principles and the assembly sequence selected should minimize the thruput time. The VRL discipline is the only meaningful launching discipline and the only influence that 'restricted operator' movement has, is to lengthen the thruput time, i.e. it does not affect the assembly sequence in any way.

6. HYBRIDS—ASSEMBLY LINES HAVING BOTH OPEN AND CLOSED STATIONS

Mixing open and closed stations in assembly lines does occur in practice. Most stations in such lines have open boundaries except those few that are closed for technological or safety considerations. However, only the 'Product Fixed' category of lines are affected, i.e. Cases (a) and (b).

Hybrid lines can be solved on a cumulative basis, applying the models (a) and (b) to the appropriate places on the assembly line. Consider for example, a 12 station assembly line having the following composition of open and closed stations:

Station:	(1)(2)(3)(4]	[5]	[6)(7)(8]	[9]	[10)(11)(12]
Case:	(b)	(a)	(b)	(a)	(b)

Key: [ ] closed station. [ ) closed upstream, open downstream. ( ) open station. ( ] open upstream, closed downstream.

Then the lower bound of the overall assembly line, $(SLONG)_{LB}$ would comprise the following terms:

$$(SLONG)_{LB} = (LB)_{1\to4} + (LB)_5 + (LB)_{6\to8} + (LB)_9 + (LB)_{10\to12}$$

The approach proposed in problem (b) should be used for generating several solution sequences which are evaluated to obtain the least cost solution.

Choosing the launching discipline is not so simple. Keeping in mind that hybrids apply only to the 'products fixed' category, it is more likely that industrial practice would lean towards employing the FRL discipline, whereas it is suggested that the VRL discipline could produce more efficient solutions. However, the ultimate choice is more likely to be made on the basis of a cost analysis of alternative conveyor systems.

7. DISCUSSIONS AND CONCLUSIONS

It is hoped that this paper has succeeded in incorporating all mixed-model assembly line types likely to be found in practice. Illustrations of the various assembly line categories were given and their behaviour analysed. It was shown that two design criteria exist in mixed-model sequencing problems:

- to find the sequence that minimizes the overall facility length
- to find the sequence that minimizes the thruput time.

Although the analysis for the different catergories have led to possible solution methods, more importantly, it has provided an insight as to what design decisions ought to be made in the first place. The designer should initially enquire whether a 'Product Movable' arrangement (models (c) and (d)) is feasible, since the resulting layout is likely to be far less expensive than for the 'Product Fixed' arrangement. Also, the possibility of a buffer stock occurring with 'Product Movable' lines means that a shorter cycle time can be used.

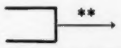
However, the greatest advantage for using 'Product Movable' lines lie in its flexibility to cope with changes in production, since the layout is *independent* of the sequence. This is not the case with 'Product Fixed' lines, which requires a 'new solution' to be produced for each production change.

If the product *must* be fixed to the conveyor, then as many stations as possible should have open boundaries, with the VRL discipline selected as the preferred launching system. These design factors are summarised in Table 2.

Research efforts are needed in two main areas (marked ** in the table):

- In formalising a method for solving the sequencing problem in 'Product Fixed–Open Station' lines (including hybrids)

TABLE 2. SUMMARY OF DESIGN FACTORS

Station type (case)	Products fixed: Closed (a)	Products fixed: Open (b)	Products movable: Closed (c)	Products movable: Open (d)
Launching system	Fixed rate Launching(FRL) or Variable Rate Launching(VRL)	FRL or VRL†	VRL†	VRL†
Movement across boundaries	—	Restricted or unrestricted	—	Restricted or unrestricted†
Sequencing criteria	Min. line length	Min. line length and min. thruput time	Min. thruput time	Min. thruput time
Relative Facility Length (RFL)	$(RFL)_a$	$(RFL)_b < (RFL)_a$	$(RFL)_c < (RFL)_b$	$(RFL)_d = (RFL)_c$
Relative Thruput Time (TP)	TP_a	$TP_b < TP_a$	$TP_c < TP_b$	$TP_d < TP_c$ (marginally)
Flexibility in coping with Production changes	None	None	Maximum	Maximum
Preference based on Total Cost	2nd Preference (closed stations used only when essential)		1st Preference (closed stations used only when essential)	
Future research effort	*	**		

† Preferred

- In developing inexpensive means for transposing 'Product Fixed' designs into the 'Product Movable' mode.

Of less urgency, though still important, is the search for a more efficient solution to the 'Product Fixed–Closed Station' category (marked * in the table). Attention could also be directed towards tackling problems in which the overall line length is given, as input data. This problem is only relevant to the 'Products Fixed–Open Stations' category (case (b)) and to hybrid lines (case (a) would be insolvable and cases (c) and (d) are not affected). It is hoped that this paper enhances the contribution that industrial engineering practitioners can make to designers of mixed-model assembly lines.

REFERENCES

1. DAR-EL EM & COTHER RF (1975) Assembly line sequencing for model-mix. *Int. J. Prod. Res.* **13**(5). 463–477.
2. DAR-EL EM & CUCUI S (1977) Optimal mixed-model sequencing for balanced assembly lines. *Omega* **5**(3). 333–342.
3. DAR-EL EM & NADIVI A (1977) A mixed-model sequencing application. *Technion-III*. Operations Research, Statistics and Economics Mineograph Series No. 193.
4. KILBRIDGE MD & WESTER L (1963) The assembly line model-mix sequencing problem. *Proc. of the Third Int. Conf. Op. Res.* English Universities Press; (Paris: Dunod Editeur).
5. THOMOPOULOS NT (1967) Line balancing sequencing for mixed-model assembly. *Mgmt Sci.* **14**(2). 59–75.
6. THOMOPOULOS NT (1970) Mixed-model line balancing with smooth station assignments. *Mgmt Sci.* **16**(9). 593–603.
7. WILD R (1972) *Mass Production Management*. John Wiley, Chichester, UK.

ADDRESS FOR CORRESPONDENCE: *Professor Ezey M Dar-El, Technion, Israel Institute of Technology, Haifa, Israel.*

APPENDIX

The following is a brief description of the Dar-El–Cother method for determining the assembly sequence in mixed-model problems.

The method starts by making the length of each station equal to its lower bound. Products are then sequenced in a specific manner which always aims at minimizing the sum of station length increments.The lower bound $(LB)_i$

Vol. 6 1978

for each station i is found as follows:
let

$\bar{t}_i$	be the largest task time allocated to station i.
t_i	the smallest task time allocated to station i.
γ	the FRL period.
T_i	the total work allocated to station i.
P	the total number of products to be assembled over the shift (or planned period).

Then three lower bounds can be defined:

$$(\text{lb})_1 = 2\gamma - t_i \tag{A1}$$

$$(\text{lb})_2 = \bar{t}_i \tag{A2}$$

$$(\text{lb})_3 = t_i + |T_i - \gamma \cdot p| \tag{A3}$$

$$\text{Then } (\text{LB})_i = \max \{(2\gamma - t_i), \bar{t}_i, (t_i + |T_i - \gamma p|)\} \tag{A4}$$

$(\text{SLONG})_{\text{LB}}$, the lower bound for the overall assembly line length is then given by:

$$(\text{SLONG})_{\text{LB}} = \sum_{i=1}^{L} (\text{LB})_i. \tag{A5}$$

The third lower bound $(\text{lb})_3$, only applies when $T_i \neq \gamma P$ (Dar-El and Nadivi [3]) implying that the starting position of the operator in station i *does not* coincide with his finishing position (when work on all P products are completed).

Selecting candidates for the sequence depends on a 'Selection Heuristic' which is briefly described below:

If $\mathbf{M} = (M_1 : M_2 : \ldots : M_k)$ is the 'proportional vector' between models 1, 2, ..., k that are to be assembled, then the model selected for the next place on the sequence should aim at maintaining as near as possible, the vector $\mathbf{M}$ among the pool of products yet to be allocated.

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 325–330
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0325$02.00/0

The Project Coordinator's Problem

LF McGINNIS

Georgia Institute of Technology, USA

HLW NUTTLE

North Carolina State University, USA

(*Received February* 1978)

A project coordinator has a set of activities with known durations and values which are available for scheduling within a given horizon. There may be both precedence and corequisite restrictions (two activities are corequisites if scheduling one implies the other must be scheduled). The project coordinator's problem is to select the activities to be scheduled, and their start times, so as to maximize the net value of the project when there is no penalty for failing to schedule an activity. An efficient solution procedure is developed and illustrated and several generalizations are discussed.

INTRODUCTION

A PROJECT coordinator has a set of activities available for possible scheduling within a fixed planning horizon. The execution time, or duration, of each activity is known, and there is a value (possibly negative) which accrues if and only if the activity is scheduled and completed within the horizon. Not all activities must be scheduled in the horizon; however, there is associated with each activity, two subsets of activities (possibly empty). One subset consists of activities whose completion is a *prerequisite*, i.e. the activity in question may not begin until its prerequisite activities are completed. The other subset consists of activities which are *corequisites*, i.e. the activity in question may be scheduled if and only if its corequisite activities are also scheduled. There is no penalty for failing to schedule an activity.

The *project coordinator's problem* is to select the activities to be scheduled and determine their start times so that the prerequisite and corequisite restrictions are satisfied, no activity is in progress beyond the planning horizon, and the net value of the project is maximized. In the paper, we present a mathematical programming formulation of this optimization problem and an efficient method for obtaining an optimal solution.

MODEL FORMULATION

Define the following notation:

N	$= \{1, 2, \ldots, n\}$, the set of available activities;
$j \in N$	= an arbitrary activity;
$P(j)$	= the set of prerequisite activities for j;
$C(j)$	= the set of corequisite activities for j;
d_j	= the duration of activity j;
v_j	= the value of activity j;
t_j	= the start time for activity j, if j is scheduled;
T	= the length of the planning horizon;
y_j	= a 0–1 variable, $y_j = 1$ if j is scheduled, 0 otherwise.

Using this notation, the project coordinator's problem can be formulated as the following mathematical programming problem:

$$\text{PCP: maximize} \sum_{j=1}^{n} v_j y_j$$

subject to

$$\begin{aligned}
& t_j y_j - t_i y_i \geq d_i && i \in P(j), j \in N \\
& t_j y_j \leq T - d_j && j \in N \\
& y_j - y_i = 0 && i \in C(j), j \in N \\
& t_j \geq 0 && j \in N \\
& y_j = 0, 1 && j \in N
\end{aligned}$$

To simplify the development we shall assume that no redundant precedence relationships are

incorporated in the sets $P(j)$, i.e. if $i \in P(j)$ and $j \in P(k)$ then $i \notin P(k)$. At the same time, a reference to "all predecessors of j" will be understood to mean all activities i such that a sequence of precedence relationships exists linking i and j.

PCP, as stated, has both continuous and 0–1 variables and quadratic constraints. Fortunately, some fairly obvious simplifications permit the problem to be decomposed so that the selection and scheduling questions can be addressed separately and solved by computationally efficient methods.

SIMPLIFICATIONS

The prerequisite sets define a precedence graph with the activities as nodes. Also the value, v_j, obtained by selecting activity j does not depend on the start or completion time. Therefore, any set of t_j values consistent with the prerequisite restrictions will be optimal for a given set of optimal y_j values. Based on these observations, we may, without loss of generality, invoke the following simplification.

To the implied precedence graph, add a dummy activity, 0, with $v_0 = 0$ and define a precedence $(0, j)$ for every $j \in N$ such that $P(j) = \emptyset$. Let $t_0 = 0$ and let t_j, $j \in N$, be the CPM early start time, i.e.

$$t_j = \max_{i \in P(j)} \{t_i + d_i\}, j \in N$$

Any activity $j \in N$ for which $t_j + d_j > T$ may be dropped from consideration, since it cannot possibly be scheduled and completed within the given horizon. Suppose j is to be dropped. If $j \in C(k)$ for some activity k, then, clearly, k must also be dropped, along with all activities $i \in N$ such that k is a predecessor of i. Eliminate as many activities as possible in this manner.

If there is an activity, j, with $v_j \geq 0$, j is not a corequisite for any other activity, and for all i which precede j, $v_i \geq 0$ and i is not a corequisite for any other activity, then a further reduction is possible. In this case, it is clearly optimal to schedule j and all its predecessors. Not doing so will result in a smaller total value for the project, and scheduling these activities does not 'force' a decision on any other activities. Activity j and all its predecessors may also be dropped from consideration at this point.

If there exists a pair of activities i and j such that i is a predecessor of j, and i and j are corequisite, then the following reduction is possible. Consider the set, S, of activities, where S contains i, j, and all activities k such that i is a predecessor of k and k is a predecessor of j. Either all activities in S are scheduled or none of them are scheduled. Therefore, all activities in S may be combined and represented as a single activity (whose value is the sum of all v_j for $j \in S$) for the purpose of activity selection.

At this point, the remaining corequisite activities may be combined and treated as a single activity with value equal to the sum of the values of the corequisite activities. There would be one such combined activity for every set of corequisite activities. Although a valid simplification, this will not significantly alter the effort required to solve the selection problem.

After simplifying and reducing the set of activities, redefine N, and renumber the remaining activities $0, 1, 2, \ldots, n$. We now have a CPM-type network defined on the activities and a set of feasible start times. The activities to be scheduled may now be determined by solving the following activity selection problem (the variables in brackets are the dual variables):

$$(P) \text{ maximize } \sum_{j=0}^{n} v_j y_j$$

subject to

$y_j - y_i \leq 0$	$i \in P(j), j \in N, P(j) \neq \emptyset$	$[w_{ij}]$
$y_j - y_i = 0$	$i \in C(j), j \in N, C(j) \neq \emptyset$	$[w_{ij}]$
$y_j = 0, 1$	$j \in N$	$[w_j]$

SOLUTION PROCEDURE

The activity selection problem as stated is a 0–1 linear programming problem. We note, however, that the transpose of the matrix of constraint coefficients satisfies the Heller–Tompkins sufficient condition for total unimodularity [6]. Thus, the problem can be solved directly as a linear programming problem, with the restrictions on y_j relaxed to $0 \leq y_j \leq 1$.

The linear programming dual to this problem is (x_j is a surplus variable):

$$(D) \text{ minimize } \sum_{j=0}^{n} w_j$$

subject to

$$-x_j + w_j - \sum_{\substack{k: \\ j \in P(k) \cup C(k)}} w_{jk} + \sum_{\substack{i: \\ i \in P(j) \cup C(j)}} w_{ij} = v_j \qquad j \in N$$

Vol. 6
1978

$x_j, w_j \geq 0$ $\quad j \in N$

$w_{ij} \geq 0$ $\quad i \in P(j)$

w_{ij} unrestricted $\quad i \in C(j)$

The dual problem can be interpreted as a minimum cost network flow problem. Define a new variable s_j, where $s_j = x_j + v_j$, require $s_j \geq v_j$, and rewrite the dual constraints as

$$-s_j + w_j - \sum_k w_{jk} + \sum_i w_{ij} = 0$$

Note that s_j may be negative if $v_j < 0$. In the dual network, there is a node for each $j \in N$ plus two additional nodes. The supersource node has an arc *to* every node corresponding to $j \in N$, with flow w_j. The supersink node has an arc *from* every node corresponding to $j \in N$ with flow s_j. There is a directed arc (i, j) for w_{ij} corresponding to $i \in P(j)$. The unrestricted w_{ij} can be handled by defining two new variables w_{ij}^+ and w_{ij}^- with corresponding directed arcs (i, j) and (j, i). Then $w_{ij} = w_{ij}^+ - w_{ij}^-$. When $v_j < 0$, s_j is unrestricted and handled in a similar fashion.

Because there are quite efficient algorithms [5] for solving minimum cost network flow problems, (D) can be readily solved for any practical number of activities. After the dual solution is obtained we can construct the corresponding primal solution by appealing to complementary slackness:

(i) $w_j > 0 \Rightarrow y_j = 1 \quad j \in N$

(ii) $w_{ij} > 0 \Rightarrow y_j = y_i \quad j \in N \qquad i \in P(j) \cdot C(j)$

(iii) $s_j > v_j \Rightarrow y_j = 0 \quad j \in N$

EXAMPLE

Consider the set of activities illustrated in Figure 1, where arcs between activities (nodes) denote precedence. Suppose also that activities 10 and 11 are corequisites. The CPM early start and finish times are given in Table 1.

Suppose the planning horizon is 20 units. Activity 13 cannot be completed within the horizon, so it is dropped from consideration.

TABLE 1. ACTIVITY START AND FINISH TIMES

Activity	1	2	3	4	5	6	7	8	9	10	11	12	13
Start	0	0	3	3	4	4	6	6	6	12	12	15	16
Finish	3	4	4	8	6	8	12	9	12	15	16	24	19

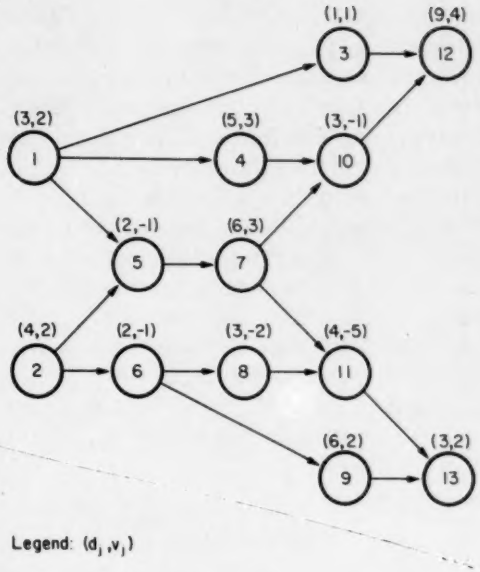


FIG. 1. *Example project.*

The network dual to the activity selection problem has the following solution:

$w_1 = 2$ $\quad w_{13} = 1$ $\quad s_{10} = 0$

$w_2 = 5$ $\quad w_{14} = 3$ $\quad s_8 = 0$

$w_{25} = 2$ $\quad s_{11} = -2$

$w_{26} = 1$ $\quad s_{13} = 2$

$w_{11,13} = 2$

All other variables are at their lower bounds. From complementary slackness, the solution to the activity selection problem is found to be:

$w_1 > 0 \Rightarrow y_1 = 1$ $\quad w_2 > 0 \Rightarrow y_2 = 1$

$w_{13} > 0 \Rightarrow y_3 = y_1 = 1$ $\quad w_{25} > 0 \Rightarrow y_5 = y_2 = 1$

$w_{14} > 0 \Rightarrow y_4 = y_1 = 1$ $\quad w_{26} > 0 \Rightarrow y_6 = y_2 = 1$

$s_{10} > v_{10} \Rightarrow y_{10} = 0$ $\quad w_{11,13} > 0 \Rightarrow y_{13} = y_{11} = 0$

$s_8 > v_8 \Rightarrow y_8 = 0$

$s_{11} > v_{11} \Rightarrow y_{11} = 0$

$w_{57} > 0 \Rightarrow y_7 = y_5 = 1$

$w_{69} > 0 \Rightarrow y_9 = y_6 = 1$

The optimal value of the project is 7, given by

$$v_1 + v_2 + v_3 + v_4 + v_5 + v_6 + v_7 + v_9.$$

DISCUSSION

Although formulated and solved independently, the project coordinator's problem is related to Rhys's [14] "selection problem of shared fixed costs." It is, in fact, a generalization of the shared fixed costs problem, since the latter did not consider the scheduling question, nor allow for corequisite activities or pre-

cedence relationships other than those characterized by a pibartite network. Balinski [1] solves the selection problem by finding a minimum weight cut in an associated graph. Picard [12] extends the results of [1] and [14] to arbitrary precedence networks.

In addition to the applications described by Rhys, Murchland [8], and Picard, the project coordinator's problem arises naturally in several contexts. In bidding interrelated projects or selecting jobs (in a job shop context) it permits optimization and sensitivity analysis of scheduling and selection decisions. It could be used as a first step in curriculum design or course content selection, because it explicitly addresses the time factor.

EXTENSIONS

A number of generalizations of the basic model are possible. Unfortunately, most extensions lead to problems which are computationally much more difficult to solve. In the following section, several generalizations are described and a brief analysis of their impact on the model is presented.

Mutually exclusive activities

Two activities may be related other than by the precedence or corequisite requirements described earlier. One such relationship is mutual exclusivity, i.e. two or more activities are mutually exclusive if the selection of any one activity precludes the selection of any of the others. In terms of the decision variables, y_j, this restriction can be written

$$\sum_{j \in M} y_j \leq 1$$

where M is the set of mutually exclusive activities. (Note that no activity should be a successor of two mutually exclusive activities, since it could *never* be selected due to one of its predecessors not being selected.) If this constraint is added to the activity selection problem, the transpose of the coefficient matrix no longer satisfies the Heller–Tompkins conditions for total unimodularity. Moreover, an attempt to formulate the network flow dual fails (except in some special cases described in [14]) because conservation of flow is violated. (A dummy source is defined for each set, M, but the *same* flow must appear on each arc to a node $j \in M$.

With mutual exclusivity relationships included, the linear programming relaxation of the activity selection problem will no longer guarantee integer solutions. Thus, the activity selection problem becomes a true integer programming problem, albeit one with a very special structure.

Alternative activities

The case of mutually exclusive activities is distinct from the case of alternative activities. In the latter situation there are two or more ways to achieve a given end, each having its own time and resource requirements and cost. The alternative activities may have exactly the same predecessors and successors, in which case the selection of exactly one of the alternatives is required for the selection of any successors.

Inclusion of alternative activities in the project coordinator's problem presents some difficulties, since the selection and scheduling aspects can no longer be treated independently. In the case where there are no interactions between distinct sets of alternative activities, the following approach could be used. Solve the problem assuming the shortest duration and largest profit (smallest cost) from the alternative activities. If this dummy activity is not selected, then none of the alternatives will be selected. If the dummy activity is selected, it is still not guaranteed that one of the alternatives will be selected.

This can be ascertained by solving a reduced problem involving only the alternative activity and its successors. A dummy activity is created for each successor of the alternative set so that the schedule for the reduced problem will be feasible for the total problem. The reduced problem is then solved for each of the alternatives, and the best of these solutions incorporated in the solution for the total problem.

Mutually inclusive activities

A third type of relationship between activities, designated "mutual inclusivity", requires that at least one activity from some subset must be selected. The form of this type of constraint in the activity selection problem is very similar to that of the mutual exclusivity constraint, *viz.*,

$$\sum_{j \in M} y_j \geq 1$$

where M is the set of mutually inclusive activities. The analysis of this extension follows that for mutually exclusive activities, i.e. it leads to a true integer programming problem.

Budget constraint

It is reasonable to postulate an upper limit on investment during the planning horizon. If the maximum allowable investment is $\$B$, the following constraint should be added to the activity selection problem:

$$\sum_{j:v_j<0} -v_j y_j \leq B$$

As with the mutual exclusivity and mutual inclusivity, a budget constraint violates the sufficient conditions for unimodularity, and makes it impossible to formulate a network flow dual to the activity selection problem.

An alternative form of the budget constrained problem is to assume an investment cost, c_j, for each project and require that

$$\sum_j c_j y_j \leq B$$

This allows for situations where a project returns part but not all of its investment. With this type of budget restriction, the activity selection problem begins to resemble many models put forth for R & D project selection (see [3] for a survey of such models and [7] for a discussion of implementation). There are two primary differences: R & D project selection models generally do not include precedence relationships; and the project coordinator's problem implicitly treats the planning horizon as one "time period." Attempts to generalize the project coordinator's problem to allow for multiple time periods (hence multiple budget constraints) within a given planning horizon will be difficult. Doing so introduces an interaction between the scheduling and selection aspects of the problem.

It is noted in passing that with the second form of the budget constraint, the activity selection problem resembles a 0–1 knapsack problem with additional precedence constraints. Since there are good codes for solving the 0–1 knapsack problem [9], a solution procedure based on Lagrangian relaxation [4, 10] of the precedence constraints is intuitively appealing and appears to have promise of computational feasibility.

Line of credit constraint

A generalization of the budget constraint is a line of credit constraint. Such a constraint requires that at any point in time the cumulative net investment (investment less return) from the set of projects is not greater than some specified amount. A line of credit constraint cannot be explicitly formulated in the model presented here. The scheduling approach of Pritsker, Watters and Wolfe [13] could accommodate such a model—see, e.g. the work of Penz [11]. A line of credit constraint is a tremendously complicating factor since it also introduces an interaction between activity scheduling and activity selection.

Resource constraints

There are two types of resource constraints to consider—limits on total resource usage and limits on the rate of resource usage. The former is simply a budget constraint where the budget is in terms of resources other than dollars. The latter leads to a variation of the resource constrained project scheduling problem. The basic model of Pritsker, Watters and Wolfe could be generalized in a manner similar to Doersch and Patterson's [2] model for maximizing the present value of a set of activities.

CONCLUSION

The project coordinator's problem is one of activity selection and scheduling. Although initially formulated as a 0–1 programming problem with quadratic constraints, its special structure leads to a very efficient solution procedure. There are many applications for the project coordinator's problem and its several extensions. Unfortunately the extensions are generally much more difficult to solve than the original problem. The addition of a budget constraint to the project coordinator's problem emphasizes its relationship to R & D project selection models. This particular extension is currently the most promising avenue for further research.

REFERENCES

1. Balinski ML (1970) On a selection problem. *Mgmt Sci.* **17**(3), 230–231.
2. Doersch RH & Patterson JH (1977) Scheduling a project to maximize its present value: a zero-one programming approach. *Mgmt Sci.* **23**(8), 882–889.

3. GEAR AE, LOCKETT AG & PEARSON AW (1971) Analysis of some portfolio selection models for R & D. *IEEE Trans. on Engng Mgmt* EM-**18**(2), 66–76.
4. GEOFFRION AM (1972) Lagrangian relaxation for integer programming. *Math. Prog. Study* **2**, 82–114.
5. GLOVER F, KARNEY D, KLINGMAN D & NAPIER A (1974) A computation study on start procedures, basis change criteria, and solution algorithms for transportation problems. *Mgmt Sci.* **20**(5), 793–813.
6. HELLER I. & TOMPKINS CB (1958) An extension of a theorem of Dantzig, in *Linear Inequalities and Related Systems* (Eds KUHN HW & TUCKER AW), Princeton University Press, Princeton, New Jersey, pp. 247–254.
7. LOCKETT AG & GEAR AE (1972) Programme selection in research and development. *Mgmt Sci* **18**(10), B575–B590.
8. MURCHLAND JD (1968) Rhys's combinatorial station selection problem. *Report LBS-TNT-68*. Transport Network Theory Unit, London Graduate School of Business Studies, London, UK.
9. NAUSS RM (1976) An efficient algorithm for the 0–1 knapsack problem. *Mgmt Sci.* **23**(1), 27–31.
10. NAUSS RM (1976) The 0–1 knapsack problem with multiple choice constraints. University of Missouri, St Louis, Missouri.
11. PENZ A (1976) A multiple resource constrained scheduling approach to large residential land developments. Paper presented at ORSA/TIMS Meeting, Miami.
12. Picard J-C (1976) Maximal closure of a graph and applications to combinatorial problems. *Mgmt Sci.* **22**(11), 1268–1272.
13. PRITSKER AAB, WATTERS LJ & WOLFE PM (1969) Multi-project scheduling with limited resources: a zero-one programming approach. *Mgmt Sci.* **16**(1), 93–109.
14. RHYS JMW (1970) A selection problem of shared fixed costs and network flows. *Mgmt Sci.* **17**(3), 200–207.

ADDRESS FOR CORRESPONDENCE: *Leon F McGinnis Esq, Assistant Professor, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia* 30332, *USA*.

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 4, pp. 331–340
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0331$02.00/0

Large Scale Communication Networks—Design and Analysis

MJ FISCHER

DA GARBIN

TC HARRIS

JE KNEPLEY

Defense Communications Agency, Reston, Virginia

(*Received July 1977; in revised form January 1978*)

The paper discusses the problem of design and analysis of large scale communication systems. An iterative process, composed of a minimum cost network design algorithm and a network performance algorithm, is presented for the solution of these problems. Computational considerations, using these algorithms, are discussed.

1. INTRODUCTION

IN A COMMUNICATION network a user assigned to one node of the network wishes to communicate with a user at another node of the network. The form of communication may be a telephone conversation or a message. Furthermore, a given user wishes to receive satisfactory service from the network. That is, he would like to be sure that a certain percentage of the time he will be able to place his telephone call successfully or that his message will reach its destination in a reasonable length of time. There are many formulations of the design and analysis problem of large scale communication systems, but we would like to discuss one formulation of interest to us. Briefly stated the formulation is as follows: "Knowing the user locations, their level of demand and desired grade of service, determine how to optimally allocate network resources, such as communication links and channels, so as to ensure that a given level of performance is met at the cheapest cost." In this paper, we discuss the methods we have used to solve this problem. Although not addressed explicitly in this paper, another related problem is the one of determining the number and locations of switching nodes.

For a given communication network one usually knows the location of the switching nodes and users of the network as well as their traffic requirements. The traffic requirements are in the form of a from–to matrix R, whose entries, r_{ij}, are the number of erlangs of traffic requiring transmission from node i to node j. Our problem is to determine where to put communications links (a direct connection between two nodes of the network) and size the links (determine the number of channels in each of the links) to meet a certain grade of service. All this is to be done at minimum cost.

There are two types of communication networks of interest to us; circuit-switched and packet-switched networks. In a circuit-switched network the customers are usually calls as in a normal telephone system. Arriving calls must find a free path or circuit in the network; no queueing is allowed. Those calls not finding a free path are lost. The grade of service for this

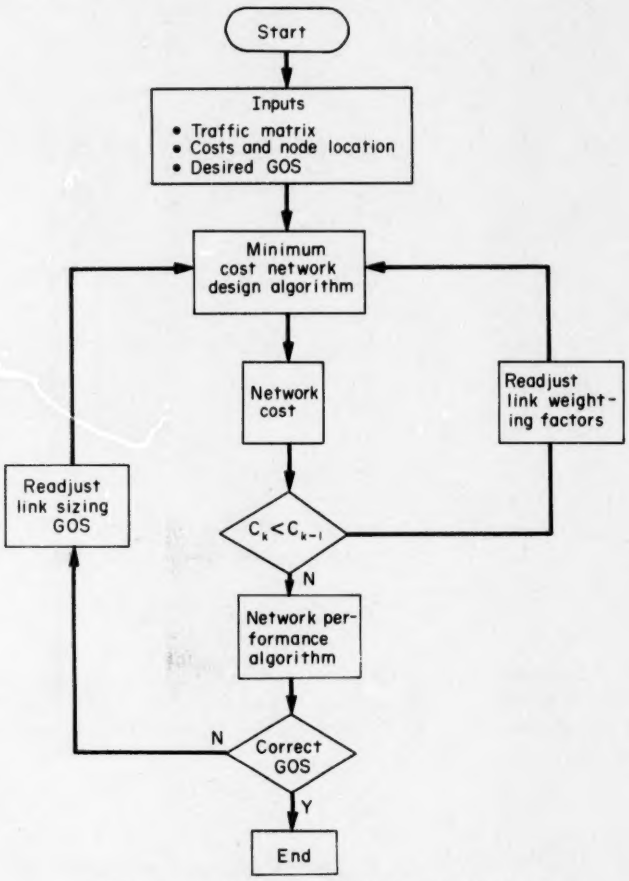


FIG. 1. *Network design and performance model.*

type of network is the probability of calls being lost. In a packet-switched network the customers are messages. Originating messages are broken up into packets by the communication equipment at the originating node and then flow through the network to their destination node where the original message is reassembled. A packet is allowed to queue at a tandem node on its path when a channel is not free. The grade of service for this network is the average packet delay.

Circuit-switched networks provide an efficient means for satisfying the communications requirements of large communities of users. This efficiency is gained through high utilization of shared resources. This quality also makes the analysis of such networks a non-trivial problem. Management of a large network requires a reliable method of predicting the network's performance under a variety of conditions. Mathematical models, as described in this paper, have been used to analyze and manage networks such as the CONUS AUTOVON system of the US Government's Department of Defense, whose annual cost approaches 100 million dollars. The design procedures described herein have been used to reconfigure the AUTOVON system with annual savings of $3 million dollars being achieved.

In the overall design and analysis of large scale communication networks, one must have the capability to predict the performance of a given network as well as to optimally configure such a network. Thus, we must have a network performance algorithm and a network optimization algorithm which are logically tied together to achieve a successful design. The overall philosophy of our method of attacking these problems is given in Fig. 1.

From this figure one sees that the overall process is an iterative one. Initially, one makes an estimate of the desired link grade of service that would result in the desired overall

Vol. 6
1978

network performance. This information is used by the minimum cost network design algorithm to determine where to put links in the network. These are sized to meet the link grade of service, and the routing table for the network is determined. The cost of the network is then computed, and if the cost at the kth iteration is less than the cost at the $(k-1)$st iteration, one returns to the minimum cost network design algorithm. If the cost is greater, the process goes into the network performance algorithm to see if the overall network performance meets the desired grade of service. If not, a new estimate is made for the link grade of service and the process is repeated.

For the circuit-switched networks, typical sizes of up to 70 nodes have to be considered. Since the overall design and analysis process is an iterative one, it is obvious that the optimization algorithm and performance algorithm have to be extremely fast. The use of the computer is required in the overall process. The design and analysis of large scale communication systems draws on three areas of interest.

a. Optimization Theory

b. Queueing Networks

c. Computer Science

Perhaps the first published work in the area of design and analysis of large communication networks was by Kleinrock [13]. He considered a store and forward communication network which is similar to the packet-switched network described above, except that messages are not packetized at the originating node. For the performance algorithm he used the results of Burke [1] and Jackson [11] to model the system as a network of queues. With this as his performance algorithm, he then considered such design problems as the optimal capacity assignment of channels so as to minimize the network delay.

The literature, mainly in conference proceedings, for packet-switched networks was driven by the establishment of the ARPA computer network in 1969. This network consisted of several computers tied together with communications links throughout the United States. The basic performance model here was one of a store and forward network along the lines of Kleinrock. Many interesting design and analysis problems resulted because of the need to understand this network. For circuit-switched networks, little has been published on the overall design and analysis problem, although the literature is full of articles on just the analysis (i.e. queueing theory). The process that we use to analyze and design these two types of networks are philosophically the same, as shown in Fig. 1. The only part of the overall process that is different is the network performance algorithm. For the packet-switched network our performance algorithm models the system as a network of Markovian queues. For the circuit-switched network the performance problem is much more difficult and we use a two moment iterative procedure to predict the performance of the networks.

Sections 2 and 3 of this paper give a description of the optimization algorithm and performance algorithm that we use in our design and analysis process. A short discussion of some of the computational considerations in the process is given in section 4.

2. MINIMUM COST NETWORK DESIGN ALGORITHM

From Fig. 1 one sees that the minimum cost network design algorithm uses the traffic matrix, network costs and the desired network grade of service to develop the cheapest network that meets all the requirements. It is possible to formulate the problem as the following non-linear integer programming problem:

$$\text{given: traffic and performance requirements}$$
$$\text{minimize: network cost} = \sum c_{ij}s_{ij} \tag{1}$$

where:

s_{ij} is the capacity (number of channels) of link (i,j)

c_{ij} is the cost per channel in link (i,j)

$$\text{constraints: conservation of flow} \tag{2}$$

$$\text{network grade of service should be met} \tag{3}$$

$$s_{ij} \text{ is a non-negative integer} \tag{4}$$

This formulation is known as a non-linear capacitated flow problem. Equation (1) represents the linear cost objective function. The cost c_{ij}, is the cost of providing one channel on link (i,j); it can be a function of the length of link (i,j). The constraint given by (2) requires

that the flow into a node of the network must equal the flow out of the node.

Dropping constraint (3) the problem becomes an integer programming problem, but the size of the problem makes the solution prohibitive. In a circuit-switched network we want to consider networks with up to 70 nodes. Fortunately, a heuristic method does exist for the solution of this problem [15, 9]. The integer programming problem may be alternatively formulated as a multiterminal shortest path problem. In this case the link costs, c_{ij}, are used as lengths and the shortest (cheapest) path between every pair of switches is found. Given this result, sufficient capacity to meet each point-to-point flow requirement is provided on each path. The final network is found by superposition of all of the link capacities of the shortest paths. The mathematical basis for this formulation is given by Hu [10] and Ford [7].

The exact solution of the non-linear capacitated flow problem represents a formidable task. We present a heuristic algorithm for its solution. This algorithm utilizes, as its basic analytical technique, the shortest path solution method presented for the linear capacitated flow problem. Utilizing the shortest path method, the non-linear problem is solved as a sequence of related linear capacitated flow problems. The new problem, which is solved at each stage of the iterative process, is obtained by a sequential modification of the objective function [13]. The objective function is modified at each iteration by a technique which weights the link costs, c_{ij}, in order to obtain a new variable, $\bar{c}_{ij}$, which represents the marginal cost to increase the link capacity per unit of flow increase. The weighting function is set equal to one in the first iteration. The first iteration, therefore, starts with the solution to the linear capacitated flow problem. If point-to-point traffic exists between every switch pair and the costs are uniform throughout the network, this first solution is a fully connected network. At each subsequent iteration, a new weighting factor, w_{ij}, is computed for each link.

This weighting function is dependent on the link loading and size, and the desired network performance. It is the negative of the derivative of the link performance measure with respect to load divided by the derivative of the link performance measure with respect to the number of channels. The resultant is the rate of change with respect to capacity divided by a corresponding change in load. The smaller this number the more traffic a unit increase in capacity can handle. Forming the weighted cost element, $\bar{c}_{ij} = c_{ij} \cdot w_{ij}$, in effect stretches or shrinks the length of each link. Efficient links are shortened and inefficient links are lengthened. The subsequent solution of the shortest path problem, given the current set of $\bar{c}_{ij}$'s, will cause certain inefficient links to be dropped from the network. The length of an eliminated link is set equal to infinity for all subsequent iterations. The network thus progresses from one which is fully connected to a network of reduced connectivity (and cost) by the process of eliminating inefficient links and shifting the point-to-point loads to more efficient, cost-effective paths. Due to the change in connectivity at each iteration, the relative efficiency of remaining network links also changes due to new link loading and sizes. These new loads and sizes generate revised weighting factors which, in turn, generate a new linear capacitated flow problem. Each iteration provides the solution of the new problem by the multiterminal shortest path method. The process converges when no further cost savings are possible.

More specifically, Fig. 2 is a flow diagram of the minimum cost network design algorithm. The required inputs are:

a. Node-to-node traffic requirements,

b. Network tariff data,

c. Switch locations, and

d. Desired network performance.

The principal steps of the minimum cost network design algorithm are:

a. Determination of link cost weighting factors,

b. Development of routing table,

c. Loading of network trunk groups in accordance with the point-to-point offered load and the network routing plan,

d. Trunk group sizing, i.e. allocating capacities to each trunk group in order that network performance requirements are satisfied,

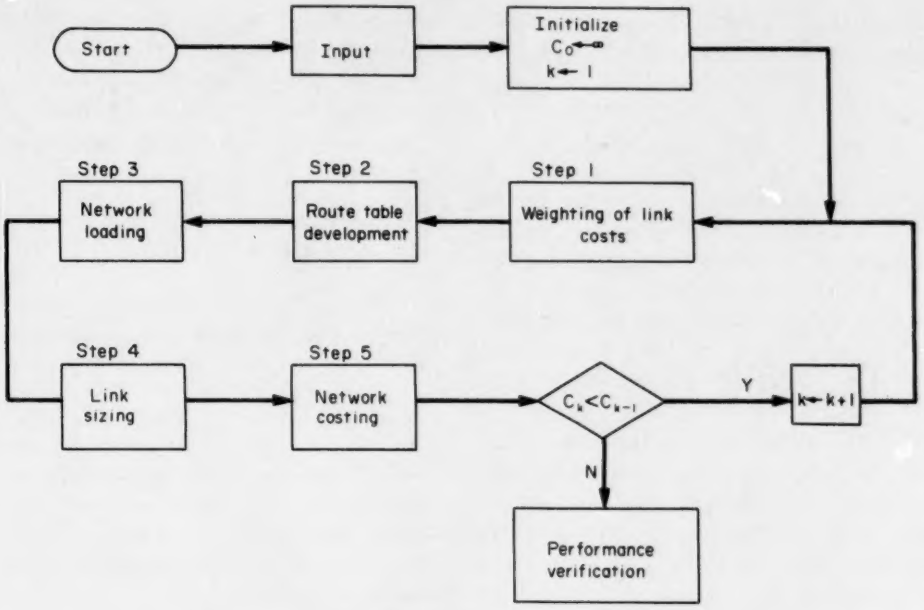


FIG. 2. *Flow chart for minimum cost network design algorithm.*

e. Computation of network cost, and

f. Check for algorithm convergence.

The output is the specification of a network; i.e. links, number of channels in each link, routing table, etc., which is fed to the performance algorithm for final verification of performance.

From the desired network grade of service an initial estimate is made for what the link grade of service ought to be in order to meet the overall network grade of service. At each iteration the initial step of the algorithm is the weighting of the link costs. Weighting factors are computed at every iteration of the algorithm and they convert the actual network link cost to marginal costs. It is well known [3] that for standard Markovian loss systems a large number of servers (or channels) are more efficient than smaller ones. Thus, we would like to readjust the cost of links (which become distances in the next step of the algorithm) so as to make efficient links shorter and inefficient ones longer.

A matrix of weighted links costs, $\overline{C}$, is formed by multiplying each link cost element by its associated weighting factor. This matrix of marginal costs provides the input to the next step of the algorithm, route table development.

Derivation of the routing plan for the minimum cost network is based on the cheapest (shortest) paths through the network. Using Floyd's shortest path algorithm, the shortest path between every node pair in the network is determined. The input to Floyd's algorithm is the marginal cost matrix, $\overline{C}$. Each element of this matrix, $\bar{c}_{ij}$, is the weighted cost per unit capacity in the trunk group connecting switches i and j. The dimension of a cost element, $\bar{c}_{ij}$, is cost per unit of offered traffic (normally expressed in terms of dollars per erlang), and $\bar{c}_{ij}$ represents the marginal cost of carrying a unit of traffic on link (i,j).

The solution of the multiterminal shortest path problem, given the input matrix, $\overline{C}$, provides two outputs:

a. A matrix comprising the marginal costs over the least cost paths for every node pair in the network, and

b. A second matrix which provides the routing for every origin-destination pair in the network.

The second output is utilized in the next step of the algorithm to distribute the point-to-point offered loads throughout the network.

The next task of the algorithm is applying the mean point-to-point offered load to the

network links in accordance with the routing plan developed in the previous step. The method used to load the trunk groups is straightforward. For the packet-switched network, only primary path routing is considered (mainly because customers are queued); thus, the offered load to each link of the network is found by distributing the originating load given in the traffic matrix along the primary paths of the network.

For the circuit-switched network customers may be blocked on a link, and since there is no queueing, they seek alternate paths in the network to the destination. For this type of network the offered load to the links of the network is found by using the routing table with alternate routing to distribute the originating traffic throughout the network via the design link blocking probability. For either type of network the offered load to each of the links of the network is determined.

The next step of the algorithm is the sizing of the links. Depending on the type of network the minimum number of channels for each link, s_{ij}, is found that ensures that the given level of link performance is met. The number of channels on each of the links is multiplied by the cost (original), c_{ij}, and summed over all links of the network to arrive at a total network cost for this iteration, c_k. If this cost is cheaper than the last iteration, the algorithm recomputes the weighting factors for another iteration. Otherwise, it ends and the previous iteration's network is fed into the performance algorithm to see if the network meets the desired network grade of service. If it does, the process stops; otherwise it updates the estimate of the required link grade of service and starts over again.

3. NETWORK PERFORMANCE ALGORITHM

The output of the minimum cost network design algorithm is the definition of a network (links, the number of chanels in each link, routing tables, etc.) that is supposed to meet the desired grade of service at minimum cost. The next step in the overall design and analysis process is to check and see if this network does indeed meet the desired grade of service.

We could use a simulation model to see if the network yields the correct grade of service, but this would be very time consuming and so we have approached the problem by developing two analytic network performance algorithms, one for each type of network.

The basic assumptions for these algorithms are:

a. Traffic originates at the nodes of the network as a Poisson process.

b. The length of time required to serve a customer can be described by an exponential distribution.

c. The network is in steady state.

d. For the circuit-switched network the mean and variance of the offered load are sufficient to describe the behavior of a customer on the links. For the packet-switched network the mean is sufficient.

e. The behavior of customers on different links of the network is statistically independent.

f. For the packet-switched networks, primary path routing is used (optimal bifurcated routing is also modeled) and customers not finding a free channel on a link are queued and served on a first-come, first-served basis. There are infinite buffers at each node.

g. Congestion in the network is only caused by the availability of channels on the links.

From these assumptions it follows that the packet-switched network may be treated as a tandem string of Markovian delay queueing systems (see [1, 11, 13]). However, there still is one interesting point to raise; the length of time to serve a given packet is the same for every link it traverses, even though, initially, this time may be exponentially distributed. That is, a given packet takes the same time to be served by different links of the network since its length is fixed. This phenomenon was considered in [13] and the results indicate that the assumption of a given packet selecting a new exponentially distributed length at each node of the network it traverses gives reasonably accurate results.

For the circuit-switched network the problem is much more difficult. No queueing is

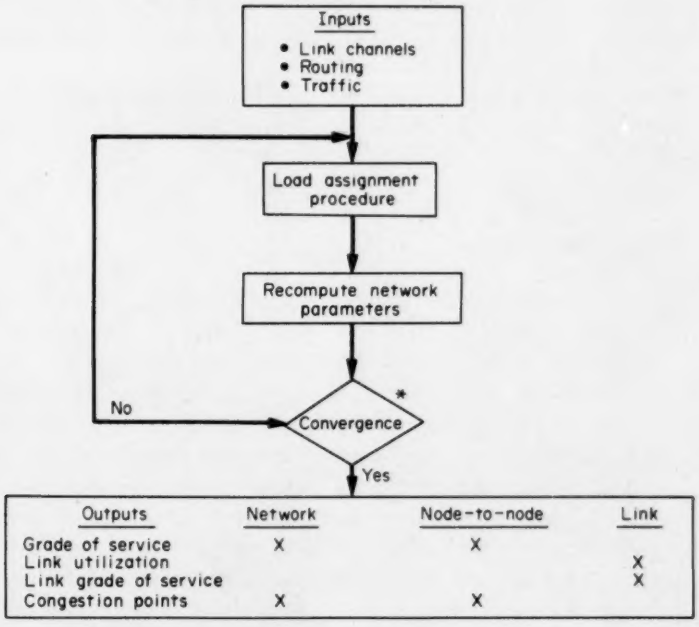
Vol. 6
1978

allowed and customers may traverse several paths in the network until they find a free one. One then is forced to consider a network of queues where at each queue no waiting is allowed and customers may have several paths to their destination from the same origin. The nature and scope of the theoretical problems one encounters are indeed challenging. It appears that no simple solution to the problem exists as in the case of the packet-switched network.

The basic problem is that the output process from a link is not Poisson and is not simply described. This is further complicated by the fact that on a given link there are three types of traffic trying to use it: originating, tandem and overflow. The originating traffic is that type of traffic that has just come into the network, the tandem has been carried by another link of the network and the overflow was blocked and is trying to find an alternate path to the destination. The arrival processes for each of these are different and not Poisson for the tandem and overflow [3]. Thus, each type of customer sees a different blocking probability [16]; and the evaluation of this probability is not straightforward [4, 16]. Thus, one is forced to consider some sort of procedure which approximates the behavior of the network. We have adapted an iterative procedure, first introduced by Katz [12], that uses the mean and variance of the offered load to describe the behavior of a customer on a link.

For both types of networks the structure of the algorithm is basically as shown in Fig. 3. The basic difference between the two network algorithms is that only one pass through the algorithm is required for a packet-switched network. This follows from the assumption that only primary path routing is used, and because of this the output processes are Poisson. So, the algorithm converges in one iteration. As discussed above, the situation with the circuit-switched networks is different and one has to iteratively approximate the magnitude of the arrival processes of each type of traffic on the links. Once these have been determined some sort of approximation is required to determine the different blocking probabilities for each arrival process.

Some of the possible outputs of the algorithm are also shown in Fig. 3. For packet-switched networks (circuit-switch) the grade of service is expected delay (probability of calls being lost).



*Not required for packet-switched networks

FIG. 3. *Network performance algorithm.*

Since the application to packet-switch networks is straightforward (we earlier discussed this point), only a discussion of the circuit-switched algorithm is given. This algorithm was first introduced by Katz [12] and has been pursued in detail by the authors. For a description of the initial implementation of the algorithm see Fischer & Knepley [5].

The performance algorithm is an iterative procedure which is based on the assumption that the traffic flow in the network can be adequately described by the mean and variance of its distribution. This two parameter characterization accounts for the non-Poisson nature of the traffic which overflows or is carried by a link.

At each iteration, a two-step procedure is implemented by the algorithm. The first step is a load assignment procedure. During this step the offered traffic is distributed over the network (one source-destination pair at a time) using the alternate routing table and the link blocking probabilities calculated at the last iteration. Initially, these probabilities are set to zero. By considering all source-destination pairs of the network, the offered load to each link of the network and the total traffic lost for all pairs are accumulated during the load assignment procedure.

The next step in the current iteration is to compute new estimates of the link blocking probabilities and the mean and variance of the overflow processes. From the load assignment procedure we know the mean and variance of the offered load to each link of the network. Based on the relationship between these parameters three methods are used to determine blocking probabilities and moments of overflow processes. If the mean and variance are equal, then we assume the offered load is being generated by a Poisson process and Erlang's Loss Formula is used to give an expression for the blocking probability. For the case where the mean is less than the variance the process is basically one which has overflowed another link [3] and Wilkenson's Equivalent Random Technique [17] is used to obtain the desired results. For the case where the mean is greater than the variance, the Smoother Than Random Technique [12] is used.

If the new estimates are not sufficiently close to the computed values of the previous iteration, the load assignment procedure is entered again using the new values of link blocking. The algorithm converges when the link blocking probabilities computed from a set of offered loads are the same as those used to generate those loads. Typically, convergence requires from seven to ten iterations. In the next section we discuss some of the computational considerations and relative accuracies of the method presented in the first three sections.

4. COMPUTATIONAL CONSIDERATIONS

In the previous sections we have presented an algorithm for designing and analyzing a circuit or packet-switched network to meet a given level of performance. Probably the most constraining feature of the problems which these algorithms were developed to solve was their size. In typical networks there are as many as seventy nodes to consider and thus thousands of variables. In this section we discuss some of our experiences in using the algorithm. Three areas are considered:

a. Optimality of the minimum cost network design algorithm.

b. Accuracy of the network performance algorithms.

c. Computer run time and storage requirements of the overall process.

The minimum cost network design algorithm presented in section 2 is a heuristic procedure that generates a local optimum for the problem. The question as to the difference in cost of the network designed with the procedure and the cost of the global optimum network is an interesting one. It has been considered to some extent in a paper by Cantor and Gerla [2] and for some special cases by Frank and Chou [8].

In their paper, Frank and Chou give an heuristic procedure for developing routing tables and their results are compared to optimum routing procedures. They consider a packet-switched type of network and formulate the design problem along the lines of section 2. Their objective function was to minimize the total delay in the network. By approximating the objective function by straight line segments they transformed their problem into a linear programming problem. They then compared

TABLE 1. NETWORK GRADE OF SERVICE COMPARISONS: ANALYTIC MODELS VS SIMULATION

	Simulation	Analytic (Single moment)	Analytic (Dual moment)
Design load	0.130	0.118	0.116
Load + 30%	0.290	0.300	0.298
Load − 10%	0.045	0.043	0.049

the results of the heuristic procedure with the ones obtained via the solution of the linear programming problem. Their results indicate that their heuristic procedure was almost as good as the optimal solution.

Although our design algorithm differs from theirs we have found from experience that there are many network configurations that give a nearly optimal solution to the problem. Many different heuristic methods can be used to get one of these solutions, ours being one such method. We feel that the method we use does achieve a network which is reasonably close to the optimally designed network.

As to the accuracy of the network performance algorithms, the case of packet-switched networks can be covered quickly. Because of the mathematical assumption made for this type of network, the results of Burke [1] can be used to justify the algorithm from a mathematical standpoint. Kleinrock [14] reported on his use of such models in attempting to predict the performance of the ARPA computer network. He found with proper modifications to the equations the results as compared with a simulation were remarkably good.

For the circuit-switched network performance algorithm the mathematical foundation is not as solid. We have made quite a few comparisons of our network performance algorithm with the results obtained from simulation. One set of these results is shown in Table 1. For that table the network we considered was composed of 35 nodes. Over a million calls were used by the simulation to estimate the grade of service.

Three cases were considered: the design load, overload and the underload case. The design load refers to the load for which the network was designed; the other two refer to the case where the input traffic was either increased or decreased by an appropriate amount.

We ran the network performance algorithm in two modes, single and dual moment. The single moment refers to the mode where all traffic in the network is modeled by a Poisson process. Whereas in the dual moment mode, the mean and variance is used to characterize the arrival processes. This was the mode described in section 3. It is worthwhile to point out that the single moment version does quite well.

The final computational consideration that we discuss are that of computer storage and run time constraints. As one would expect these are directly related to the number of nodes in the network. Since circuit-switched networks are usually much larger than the packet-switched networks we consider, the results we present were developed from the circuit-switched network design algorithm. For an IBM 360/155, the following graphs (Figs 4 and 5) relate the number of nodes to the required core in bytes and CPU run time in minutes.

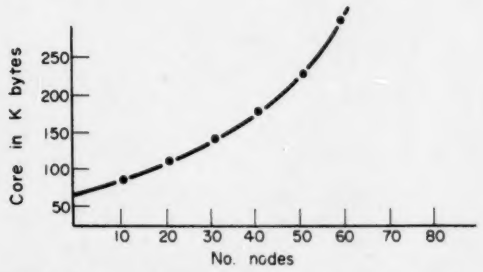


FIG. 4. *Core requirement vs number of nodes for circuit-switched network.*

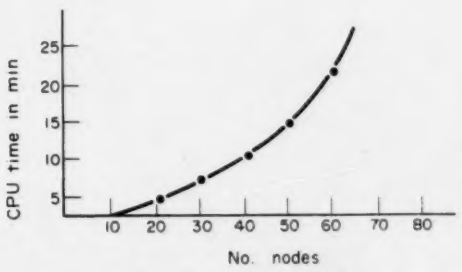


FIG. 5. *CPU time requirement vs number of nodes for circuit-switched network.*

REFERENCES

1. BURKE PJ (1956) The output of a queueing system. *Op. Res.* **4**, 699–704.
2. CANTOR DG & GERLA M (1974) Optimal routing in a packet-switched computer network. *IEEE Trans. on Computers* **C-23**(10), 1062–1069.
3. COOPER RB (1972) *Introduction to queueing theory*. Macmillan, New York.
4. DISNEY RL (1972) A matrix solution for the two server queue with overflow. *Mgmt Sci.* **19**(3), 254–265.
5. FISCHER MJ & KNEPLEY JE (1973) Circuit switched network performance Algorithm (Mod. 1). Defense Communications Engineering Center Technical Note 1–73.
6. FLOYD RW (1962) Algorithm 97: Shortest Path. *Communications of ACM* **5**(6), 345.
7. FORD LR & FULKERSON (1972) *Flows in Networks*. Princeton University Press, New Jersey.
8. FRANK H & CHOU W (1971) Routing in computer network. *Networks* **1**, 99–112.
9. GERLA M (1973) The design of store-and-forward (S/F) networks for computer communications. Ph.D. Diss., Department of Computer Science, University of California, Los Angeles.
10. HU TC (1969) *Integer Programming and Network Flows*. Addison-Wesley, Reading, Massachusetts.
11. JACKSON JR (1957) Networks of waiting lines. *Op. Res.* **5**, 518–521.
12. KATZ SS (1967) Statistical performance analysis of a switched communications network. Fifth International Teletraffic Congress, New York.
13. KLEINROCK L (1974) *Communications Nets: Stochastic Message Flow and Delay*. McGraw-Hill, New York.
14. KLEINROCK L (1970) Analytic and simulation methods in computer network design. *Proceedings of the Spring Joint Compt. Conf. AFIPS*, Vol. **36**, 569–579.
15. KNEPLEY JE (1973) Minimum cost design for circuit switched networks. Defense Communications Engineering Center Technical Note 36–73.
16. KUCZURA A (1973) Loss systems with mixed renewal and Poisson inputs. *Op. Res.* **21**, 787–796.
17. WILKENSON RI (1956) Theories for tall traffic engineering in the USA. *BSTJ* **35**(2), 421–514.

ADDRESS FOR CORRESPONDENCE: *Martin J Fischer, Esq, Defense Communications Agency, Defense Communications Engineering Center, 1860 Wiehle Avenue, Reston, Virgina 22090, USA.*

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 341–351
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0341/$02.00/0

Markov and Renewal Models for Total Manpower System

SHELBY STEWMAN

Carnegie-Mellon University, Pennsylvania, USA

(*Received November 1977; in revised form March 1978*)

This study compares the predictive utility of three stochastic models for both total manpower system and cohort personnel movement. The models are all discrete time versions, including a first order Markov chain, a Markov chain with duration of stay (semi-Markov) and a vacancy model having both renewal and Markov properties. The analysis covers a continuous 20 year period: 1950–1970 for a state police (U.S.A.) internal labor market. The simple Markov chain model is inadequate for long term cohort forecasts, but reasonably adequate for long term organizational forecasts. The semi-Markov model outperforms the simple Markov model for cohorts, but is surprisingly less accurate for the total organization. The heuristic information it portrays for the cohort is, however, quite informative. The best model for intermediate (5 year) and long term (10 year) forecasts in both cohort and organizational tests is the renewal type vacancy model. This finding is viewed as particularly important both in terms of empirical performance, which we expect can be improved due to the initial simplifying assumptions used, and in terms of further theoretical explication of the underlying causal process since internal staff flows are conceptualized as contingent on the opportunities available.

1. INTRODUCTION

THIS paper compares the predictive utility of three stochastic models for a total manpower system and a cohort sub-system. Two basic conceptualizations of mobility are represented: a Markovian push process and a renewal or job vacancy pull process. The latter model shares the perspective of a vacancy filtering throughout the system of jobs as in both [11] and [13], but is a modification of the static renewal model in [1] to include growth and allocation of new jobs. The tests cover a continuous ten-year period, 1960–1970, for a State Police (U.S.A.) internal labor market with parameters estimated from the prior decade, 1950–1960.

1.1 *Total manpower systems and cohort sub-systems*

In previous papers [9], [10] and [12] the importance of distinguishing between a total sytem and a cohort sub-system was discussed. In the present paper this importance is demonstrated.

The concept manpower system itself implies a boundary for a set of jobs. The total manpower system simply refers to the personnel flows within and across (entrances, exits) the administered set of jobs. During a given time interval, personnel enter and exit the system as it continuously operates. Thus, the population holding jobs is continually changing. Alternatively, in a cohort sub-system the population never changes. Rather, personnel flows within and across (entrances *or* exits but not both) the administered set of jobs refer to a specific individual cohort or closed population. For instance, if one is testing a model for a cohort, a sample is taken from a given point in time. Depending on whether the study is retrospective or forward from the sampling time, individual career histories are then examined for movement in prior or subsequent years in which the persons have been in the administered job system. For the forward case, during a given time interval, personnel from the original cohort may exit the job system as it continuously operates. However, whether the

job system is growing, stable, or contracting in size no entrants are analyzed by the cohort analyst. The manpower system has been artificially closed for analysis purposes. Thus, what is being considered by a cohort analyst is not the continuously operative labor market dynamics, but a subset of such dynamics. In sum, the total system model is an organizational or labor market model and the cohort model is a sub- aggregate or career model.

1.2 *Prior manpower system analyses*

Previous analyses of manpower systems have reached different conclusions about the predictive utility of a first order Markov chain model. Generally the model has demonstrated a rather high degree of predictive power (cf. [3], [5], [8], [12], [14] and [15]). Analyses in which the model has provided accurate predictions range over a variety of manpower systems—higher education, manufacturing, private firm research and development, governmental scientists, a women's military unit, a civilian manpower force which supports a national naval department and a state police organization. A different finding was recently reported concerning a national insurance company [7]. Rather than examine the types of job systems analyzed, two observations are made. All systems analyzed were internal labor markets (cf. [4, 7]). More importantly, in all cases where the analyst concluded the model's predictive power was high, the manpower involved the total system. In the dissenting case, the manpower pertained to a cohort or sub-system

Two types of comparative analyses are made in this paper. The first type pertains to comparisons across models within total systems or cohort sub-systems. For example, in the total system tests we compare a simple Markov chain, a semi-Markov and a renewal model. The same three models will also be compared for the cohort sub-system tests. The second type of comparison involves each model's capability in predicting both total system and cohort behavior. This second type of comparison should further clarify what range of predictions (aggregate, subaggregate) each model can be expected to handle.

2. THREE STOCHASTIC MODELS

The models are formulated in discrete time. In each case, the more elementary cohort model is presented first and then the inflow process is incorporated to construct a total system version.

2.1 *Markov Chain Model*

We first divide the population into k grades. Let $n_j(t)$ denote the observed number of persons in grade j at time t, $(t = 0, 1, 2, \ldots)$. The initial grade sizes $n_j(0)$ are assumed to be given. The primary task of the model is to predict the manpower system's redistribution by grade for $t > 0$. This derived *expected* number will be denoted by $\bar{n}_j(t)$.

Let the probability that an individual in grade i will move to grade j in one discrete time period be p_{ij}, and the probability of exiting to the outside from grade i be p_{io}. For this model probabilities are assumed to be stationary and therefore a time argument is unnecessary. Since there are exits from the system population,

$$\sum_{j=1}^{k} p_{ij} < 1.$$

The internal or through flows (p_{ij}'s), when multiplied by grade size at time $t-1$, generate the expected cohort distribution by grade at time t. The equation is

$$\bar{n}_j(t) = \sum_{i=1}^{k} p_{ij}\bar{n}_i(t-1) \tag{1}$$

where

$$t = 1, 2, \ldots,$$
$$j = 1, 2, \ldots, k,$$
$$\bar{n}_i(0) = n_i(0).$$

To extend Equation (1) such that it describes a total system we must include an inflow process. In order to maintain its size, the system must obviously replace all persons who leave. Similarly, if the system is to grow, not only must all leaving members be replaced, but additional recruits must enter to fill new jobs. However, in some years recruitment may be less than attrition and the job system contracts. Thus, the most general model is, as in [1] or [2],

$$\bar{n}_j(t) = \sum_{i=1}^{k} p_{ij}\bar{n}_i(t-1) + \bar{R}(t)p_{oj} \tag{2}$$

where $t = 1, 2, \ldots,$; $j = 1, \ldots, k$; $\bar{R}(t)$ is the expected number of recruits; and p_{oj} is the probability that a recruit will enter the system at grade j such that

$$\sum_{j=1}^{k} p_{oj} = 1.$$

Vol. 6
1978

Since no heterogeneity within grade level is represented by the models, the same parameter estimation procedure holds for both total system and cohort subsystem models. The internal and exit parameters have been obtained using the following maximum likelihood estimates (MLE):

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_i(t)} \tag{3}$$

$$\hat{p}_{io} = \frac{\sum_{t=1}^{T} n_{io}(t)}{\sum_{t=1}^{T} n_i(t)} \tag{4}$$

Terms:
$n_{ij}(t)$: The observed number of persons moving from grade i to grade j during the time period $(t, t+1)$; $n_i(t)$: The observed number of persons in grade i at the beginning of the time period $(t, t+1)$; $n_{io}(t)$: The observed number of persons leaving the system from grade i in the time period $(t, t+1)$.

With respect to inflow parameters, the State Police system allows no lateral entry of new recruits. Thus, $\hat{p}_{oj}$ is given by a system recruitment policy and its value is zero for all grades except the lowest where it is unity. No estimation procedure nor derived value was used for system growth.[1] The observed recruitment term $\bar{R}(t)$ was taken as given.

2.2 *Markov chain model with duration of stay*

In terms of basic equations, this model differs from the simple Markov chain model by partitioning the population into k grades *and* duration of stay per grade. Hence, it is a discrete time semi-Markov model, allowing non-stationarity of transition probabilities per grade as a function of fluctuations in duration of stay distributions. Let ${}_d\bar{n}_i(t)$ denote the expected number of persons in grade i with duration of stay d and ${}_dp_{ij}$ represent the associated transition probabilities. The ${}_dp_{ij}$ are assumed stationary. Thus, while the aggregate grade mobility is non-stationary, the underlying transition components are assumed constant. The fluctuations in the aggregate $p_{ij}(t)$ are the result of different weightings of ${}_dp_{ij}$ over time, the weights corresponding to the proportion of persons in that grade with duration d. Denote the vector of duration-grade persons $[{}_d\bar{n}_i(t)]$ as ${}_d\bar{N}(t)$. The cohort model is thus

$${}_d\bar{N}(t+1) = {}_d\bar{N}(t)\,{}_dP \tag{5}$$

and the total system model is

$${}_dN(t+1) = {}_dN(t)\,{}_dP + \bar{R}(t)\,{}_1\mathbf{P}_o \tag{6}$$

where ${}_1\mathbf{P}_o$ is a column vector of zeros except for the last k elements, which allocate the recruits, $R(t)$ to grades with duration 1.

Parameters may be estimated using the same formulae as for the simple Markov chain, except that for the present model all terms are duration specific. Further, since it was observed in Section 2.1 that all recruits enter at the bottom grade, all values of ${}_1\mathbf{P}_o$ will be zero except for the last element, which will be one.

2.3 *Renewal model*

This model differs from the previous two Markovian models by incorporating job vacancies as promotion opportunities. For the simple case where (1) promotion is allowed only to the immediately higher grade, (2) no jobs are abolished, (3) there is no lateral entry and (4) system size is constant, Bartholomew [1] developed the discrete time, equilibrium equation as follows

$$\bar{p}_{i,i+1}(t) = \frac{\sum_{j=i+1}^{k} \bar{n}_j(t)\hat{p}_{j,k+1}}{\bar{n}_i(t)} \tag{7}$$

where the numerator is the *cumulative* loss rate from all higher grades generating job vacancies and the denominator is the number of persons in the origin grade or the promotion pool. Loss rates in the above case are assumed Markovian and stationary. Given such loss rates and the initial staff distribution, $[n_j(0)]$, the model derives the expected promotion probabilities. Then, $p_{ii}(t) = 1-[p_{i,i+1}(t) - p_{io}]$. Since promotion rates are non-stationary, so are the non-movement rates.

Growth is introduced by taking into account new job vacancies as well as vacancies arising

[1] The growth and associated recruitment process for this system are not easily modeled. There exist cyclic expansion periods dependent on political and economic conditions. In addition, since the system has a cohort (recruit school) entry process rather than individual time point entry, there is a lag time from manpower allocation to entry, and the system itself has no precise specification procedure for predicting manpower losses. Moreover, using observed values for growth or recruitment, we may more closely examine the nature of the model's predictive errors for comparison with those of the closed system.

from attrition of staff. Hence, the equation becomes

$$\hat{p}_{i,i+1}(t) = \frac{\sum_{j=i+1}^{k} \bar{n}_j(t)\hat{p}_{j,k+1} + \bar{v}_j(t)}{\bar{n}_i(t)} \tag{8}$$

where $\bar{v}_j(t)$ is the expected number of new jobs created in grade j. Observe that vacancies arising from both attrition and growth are cumulative as they filter downward through the system.

To complete the model, there must be some mechanism for allocating new jobs to grades. In the simplest case, which will be used here, the relative grade sizes are assumed constant. Hence, if growth, $G(t)$, is taken as observed, as is recruitment, $\bar{R}(t)$, in the Markov models of Sections 2.1 and 2.2, then

$$\bar{v}_j(t) = G(t)\hat{\rho}_j \tag{9}$$

where

$$\hat{\rho}_j = \frac{\sum_t n_j(t)}{\sum_t \sum_j n_j(t)} \tag{10}$$

The cohort and total system models differ simply by the $n_j(t)$ for $t > 0$ since no recruits are taken into account in the cohort model, whereas recruits enter the vacant jobs in the lowest grade in the total system model. In other words for grade 1, in the cohort case,

$$n_1(t) = n_1(t-1)p_{11}(t-1), \tag{11}$$

whereas in the total system case

$$n_1(t) = n_1(t-1)p_{11}(t-1) + \bar{R}(t)p_{01} \tag{12}$$

where $\bar{R}(t)$ is observed when $G(t) < 0$ and is derived as follows when $G(t) > 0$:

$$\bar{R}(t) = \sum_{j=1}^{k} n_j(t)p_{j,k+1} + \bar{v}_j(t) \tag{13}$$

3. THE MANPOWER SYSTEM

3.1 *The data*

In a formal organization such as a state police, there are several internal labor markets. Two are most easily identified—police and civilian. Moreover, within the latter we may distinguish internal labour markets for clerical, business administration, computer programming, and maintenance staff. Since there is no movement between the police and any of the civilian internal labor markets, the latter will be excluded from the analysis and we will focus only on the police segment of the organization.

In the State Police system, promotion from the Sergeant rank or above was possible after a waiting period of one year from date of entrance. Since more than one move per man per time interval is not taken into account by the models, a one year time interval was chosen. The data extend from January 1, 1950 to January 1, 1970 and were collected for each year. Parameter estimates are made from the first ten years (1950–60) and the tests cover the second decade (1960–70).

3.2 *State definitions and parameter estimates*

The State Police system, organized along a para-military hierarchy, has a clearly defined authority and status structure. On the basis of numbers and substantive meaning of authority-status levels, the system's nine ranks were lumped into five grades. They are (1) COL: Colonel, Lt. Colonel, Major, Captain, Detective (Det.) Captain; (2) LT: Lt., Det. Lt.; (3) SGT: Staff Sergeant (S/Sgt.), Det. S/Sgt., Sgt., Det. Sgt., (4) CPL: Corporal, Det., and (5) TPR: Trooper.

Table 1 presents the MLE annual transition matrix for the simple Markov chain (MC) model estimated from Equations (3) and (4). The probability of promotion increases as an individual moves up the hierarchy with the exception of SGT. One's promotion chances in the SGT grade are essentially no better than a TPR's. In other words, there is a promotional bottleneck in middle level administration. However, if one does move upward from SGT, the probability for continued upward mobility is increased by a factor of two. The decrease in exit probabilities for CPL relative to TPR is no doubt a reflection of general service seniority accumulation and an initial success. Otherwise, exit probabilities are rank ordered such that higher grades have higher exit probabilities. This, it seems represents age, organizational seniority, and the system's retirement policies. Loss rate from modes other than retirement are quite small once the SGT rank has been achieved, (cf. [11]).

Table 2 presents the analogous MLE annual transition probabilities for the MC model with duration of stay. Due to grade size when partitioned by duration, only three partitions were feasible: 1–5, 6–10, 11 and over. For the TPR

Vol. 6
1978

TABLE 1. TEN YEAR MAXIMUM LIKELIHOOD ESTIMATES OF TRANSITION PROBABILITIES (1950-59)

Origin Grade	Destination grade						
	COL	LT	SGT	CPL	TPR	Out	N
COL	0.8523					0.1477	149
LT	0.1547	0.7459				0.0994	181
SGT		0.0674	0.8664	0.0012		0.0649	816
CPL			0.0935	0.8946	0.0013	0.0106	1594
TPR				0.0484	0.9279	0.0237	5436

grade, the first year is one of probation, and a different status is assigned at the end of the year. Thus for our estimates, we distinguish between two TPR grades: (1) the first year only, which is labeled duration 0 and (2) durations 1–5, 6–10, 11 and over, once probation is past.

We summarize Table 2 and compare it with Table 1 for promotions and exits as follows:

	Promoted (From)				Exit (From)				
	LT	SGT	CPL	TPR	COL	LT	SGT	CPL	TPR
D_0									0.0369
$D_{1,5}$	0.1697	0.0647	0.0586	0.0007	0.1170	0.0788	0.0239	0.0059	0.0235
$D_{6,10}$	0.0	0.0823	0.1988	0.0942	0.2609	0.6	0.1329	0.0086	0.0190
D_{11+}	0.0	0.0563	0.1594	0.2688	0.1613	0.1818	0.2535	0.1014	0.0163
AVE	0.1547	0.0674	0.0935	0.0484	0.1477	0.0994	0.0649	0.0106	0.0237

where D denotes duration and AVE the weighted average, which is the estimate of the simple MC model (Table 1). For the promotion process, the model's more detailed representation is quite informative. At the LT level, promotion occurs either in the first five years or never. The SGT rates are fairly constant (0.06–0.08) and the difference in promotion chances between SGT and LT is once again obvious. At the CPL level, we observe that $D_{6,10} > D_{1,5}$ by a factor of three, suggesting rather high promotion chances after 5 years similar to that of LT staff. Promotions from the TPR grade are virtually zero in the first 6 years and then change by 9%; then after an

TABLE 2. TEN YEAR MAXIMUM LIKELIHOOD ESTIMATES OF DURATION SPECIFIC TRANSITION PROBABILITIES (1950–59)

Origin Grade	Destination grade						
	COL	LT	SGT	CPL	TPR	Out	N
				Duration 1–5			
COL	0.8830	0	0	0	0	0.1170	94
LT	0.1697	0.7515	0	0	0	0.0788	165
SGT	0	0.0647	0.9114	0	0	0.0239	587
CPL	0	0	0.0586	0.9346	0.0008	0.0059	1178
TPR	0	0	0	0.0007	0.9757	0.0235	2761
				Duration 6–10			
COL	0.7391	0	0	0	0	0.2609	23
LT	0	0.4	0	0	0	0.6	5
SGT	0	0.0823	0.7785	0.0063	0	0.1329	158
CPL	0	0	0.1988	0.7896	0.0029	0.0086	347
TPR	0	0	0	0.0942	0.8869	0.0190	1370
				Duration 11+			
COL	0.8387	0	0	0	0	0.1613	31
LT	0	0.8182	0	0	0	0.1818	11
SGT	0	0.0563	0.6901	0	0	0.2535	71
CPL	0	0	0.1594	0.7391	0	0.1014	69
TPR	0	0	00	0.2688	0.7149	0.0163	491
				Duration 0			
TPR	0	0	0	0	0.9631	0.0369	814

TABLE 3. TRANSITION PROBABILITIES DERIVED FROM RENEWAL MODEL

	Promote				Stay				
	LT	SGT	CPL	TPR	COL	LT	SGT	CPL	TPR
1960	0.1276	0.0471	0.0550	0.0182	0.8523	0.7730	0.8880	0.9344	0.9581
1961	0.1316	0.0490	0.0582	0.0204	0.8523	0.7689	0.8861	0.9312	0.9559
1962	0.1349	0.0505	0.0608	0.0221	0.8523	0.7657	0.8846	0.9286	0.9542
1963	0.1536	0.0591	0.0755	0.0319	0.8523	0.7470	0.8760	0.9139	0.9444
1964	0.2024	0.0818	0.1143	0.0577	0.8523	0.6981	0.8533	0.8751	0.9186
1965	0.1761	0.0701	0.0945	0.0446	0.8523	0.7245	0.8650	0.8949	0.9317
1966	0.2387	0.0995	0.1450	0.0785	0.8523	0.6619	0.8356	0.8444	0.8978
1967	0.1741	0.0701	0.0946	0.0448	0.8523	0.7265	0.8650	0.8948	0.9315
1968	0.1990	0.0778	0.1079	0.0539	0.8523	0.7106	0.8573	0.8815	0.9224
1969	0.1581	0.0631	0.0827	0.0369	0.8523	0.7425	0.8720	0.9067	0.9394
HI RATE	0.2387	0.0995	0.1450	0.0785	0.8523	0.7730	0.8880	0.9344	0.9581
LO RATE	0.1276	0.0471	0.0550	0.0182	0.8523	0.6619	0.8356	0.8444	0.8978
Range of Variation	0.1111	0.0524	0.09	0.0603	0.0	0.1111	0.524	0.09	0.0603

* Equation (8).

† Residual of the promotion and exit rates, the latter of which are assumed constant with magnitudes as reported in Table 1.

additional 5 years, they increase by another 18% and are very high, with 27% of staff promoted per year. Similar large variations by grade seniority may be observed for exits from the higher grades. At TPR, however, attrition is only 4% in the first year and approximately 2% thereafter.

The renewal model requires an estimate of the relative allocation of new jobs per grade, $\hat{\rho}_j$ (Equation 10). For the period 1950–59, the relative size distribution was as follows:

COL	LT	SGT	CPL	TPR
0.0182	0.0221	0.0998	0.1950	0.6649

Using these estimates, the transition probabilities derived from the renewal model are shown in Table 3. Considerable variation in promotion rates are produced by this model between 1960 and 1970. The last row of Table 3 summarizes the range of variation: LT, 11%; SGT, 5%; CPL, 9% and TPR, 6%. The stay rates are necessarily of the same magnitude of change since we assumed constant exit rates.

[2] The observed data reported in Table 4 differ slightly from those in a previously reported test [12]. These differences stem from the fact that two distinct data sets were used in each case and from a different accounting system for the TPR grade. The small differences in all upper grades indicates the consistency of the data between personnel rosters and individual career files. The differences in the TPR grade is due to a difference in the analyst's date of allocating recruits. In [12] recruits were counted when they left recruit school and entered the field; here recruits are observed at the date they enter recruit school.

4. TESTS

In a previous test, using MLE from the entire 20 year time period (1950–70), the simple MC model was applied to the total system [12]. In the present test, the MLE are obtained from the first 10 years (1959–60) and predictions pertain to the second 10 year period (1960–70), thus having greater correspondence to actual forecast situations. Also, cohort predictions from the MC model are made here. More importantly, comparative tests between the MC model, the MC model with duration of stay and the renewal model are conducted for both total system and cohort sub-system dynamics. To date, only one total system test of the MC with duration has been reported [15] and no tests of the renewal model have been reported.

Table 4 presents the predicted and observed grade sizes for each model from 1961–1970. These data pertain to the total system.[2] Table 5 shows corresponding tests for the 1960 cohort only. In both total system and cohort tests the initial year, 1960, for subsequent predictions is taken as observed. Thereafter, all grade sizes are derived. For instance, while the 1961 prediction is based on the 1960 observed values, the 1962 forecast uses the 1961 prediction; the 1963 forecast uses the 1962 prediction and so on. Thus, a ten year forecast or test is involved for the 1970 predictions.

The first type of comparison we make is between models—for the total system and then

TABLE 4. PREDICTED AND OBSERVED GRADE DISTRIBUTIONS BY YEAR FOR THREE MODELS: TOTAL SYSTEM

Year	Model[a]	Grade COL	LT	SGT	CPL	TPR	Year	Model[a]	Grade COL	LT	SGT	CPL	TPR
1961	MC	20	24	112	231	754	1966	MC	25	33	147	282	894
	MCD	20	24	113	222	763		MCD	23	31	162	323	838
	R	19	22	106	216	777		R	23	27	127	257	916
	O	19	24	111	231	745		O	20	19	119	226	955
1962	MC	21	25	119	243	735	1967	MC	26	35	154	296	1077
	MCD	20	25	120	232	746		MCD	24	34	172	341	1013
	R	19	22	106	217	780		R	26	30	143	289	1025
	O	19	22	115	228	751		O	23	16	128	296	1040
1963	MC	22	27	126	253	726	1968	MC	28	36	161	317	1131
	MCD	21	26	130	241	726		MCD	26	36	182	355	1067
	R	19	22	107	219	787		R	27	32	151	304	1079
	O	21	19	119	225	760		O	23	22	137	335	1066
1964	MC	23	29	113	262	777	1969	MC	29	38	169	339	1220
	MCD	21	27	141	268	730		MCD	27	38	192	384	1144
	R	20	23	111	225	808		R	29	35	162	326	1154
	O	18	17	121	229	791		O	23	22	148	351	1152
1965	MC	24	31	140	271	815	1970	MC	31	40	178	363	1246
	MCD	21	29	151	302	773		MCD	29	41	205	410	1161
	R	22	25	120	243	871		R	30	36	169	338	1196
	O	20	16	122	227	886		O	24	22	157	389	1166

[a]MC: Markov Chain.
MCD: Markov Chain with Duration.
R: Renewal.
O: Observed Values.

TABLE 5. PREDICTED AND OBSERVED GRADE DISTRIBUTIONS BY YEAR FOR THREE MODELS: COHORT

Year	Model[a]	Grade COL	LT	SGT	CPL	TPR	Year	Model[a]	Grade COL	LT	SGT	CPL	TPR
1961	MC	20	24	112	231	722	1966	MC	25	32	143	253	498
	MCD	20	24	113	222	731		MCD	23	31	162	323	416
	R	19	22	106	216	745		R	23	27	127	257	522
	O	19	24	111	231	712		O	20	19	119	226	527
1962	MC	21	25	119	242	670	1967	MC	26	34	148	251	462
	MCD	20	25	120	232	680		MCD	27	34	172	340	356
	R	19	22	106	217	712		R	26	30	143	289	438
	O	19	22	115	228	685		O	23	16	128	295	389
1963	MC	22	27	126	249	622	1968	MC	27	35	152	247	429
	MCD	21	26	130	241	627		MCD	26	36	182	352	300
	R	19	22	107	219	678		R	27	32	151	304	381
	O	21	19	119	225	650		O	23	22	137	323	309
1964	MC	23	29	132	253	577	1969	MC	28	36	155	242	398
	MCD	21	27	141	268	556		MCD	27	38	192	375	233
	R	20	23	111	225	637		R	29	35	162	326	314
	O	18	17	121	229	612		O	23	22	148	328	261
1965	MC	24	31	138	254	536	1970	MC	29	37	157	236	370
	MCD	21	29	151	301	479		MCD	29	41	204	393	168
	R	22	25	120	243	575		R	30	36	169	338	264
	O	20	16	122	227	588		O	24	22	157	355	200

[a]MC: Markov Chain.
MCD: Markov Chain with Duration.
R: Renewal.
O: Observed Value.

TABLE 6. PREDICTIVE ERRORS FOR TOTAL SYSTEM (TS) AND COHORT SUB-SYSTEM (CS) OVER TIME† AND ACROSS GRADES*

		Total system					Cohort system						
		Grade					Grade					All grades	
Time	Model[a]	COL	LT	SGT	CPL	TPR	COL	LT	SGT	CPL	TPR	TS	CS
		Chi square values* over selected years per grade and across grades											
	MC	0.05	0	0.01	0	0.11	0.05	0	0.01	0	0.14	0.17	0.20
1961	MCD	0.05	0	0.04	0.36	0.32	0.05	0	0.04	0.36	0.49	0.87	0.94
	R	0	0.18	0.24	1.04	1.32	0	0.18	0.24	1.04	1.46	2.87	2.92
	MC	0.67	7.26	2.31	7.14	6.19	0.67	7.26	1.86	2.87	5.04	24.57	17.70
1965	MCD	0.05	5.83	5.57	18.63	16.52	0.05	5.83	5.57	18.19	24.80	48.53	54.44
	R	0.18	3.24	0.03	1.05	0.26	0.18	3.24	0.03	1.05	0.29	4.76	4.79
	MC	1.58	8.10	2.48	1.86	5.14	0.86	6.08	0	60.0	78.11	19.16	145.05
1970	MCD	0.86	8.80	11.24	1.08	0.02	0.86	8.80	10.83	3.67	6.1	22.0	30.26
	R	1.20	5.44	0.85	7.70	0.77	1.20	5.44	0.85	0.86	15.52	15.96	23.87
		Average yearly relative error† over time per grade										Weighted average relative error**	
	MC	0.12	0.26	0.06	0.09	0.05	0.12	0.26	0.06	0.07	0.05	0.07	0.06
1961–1965	MCD	0.05	0.14	0.09	0.11	0.05	0.05	0.24	0.09	0.11	0.08	0.07	0.09
	R	0.06	0.16	0.07	0.05	0.03	0.06	0.16	0.07	0.05	0.04	0.04	0.05
	MC	0.17	0.35	0.11	0.07	0.05	0.14	0.34	0.08	0.18	0.15	0.07	0.15
1961–1970	MCD	0.09	0.34	0.17	0.12	0.05	0.09	0.34	0.17	0.13	0.11	0.08	0.13
	R	0.11	0.27	0.09	0.07	0.02	0.11	0.27	0.09	0.05	0.09	0.04	0.08

† $\frac{|\text{Predicted} - \text{Observed}|}{\text{Predicted}}$. * $\chi^2 = \frac{(\text{Predicted} - \text{Observed})^2}{\text{Predicted}}$. ** Equation (14).

[a]MC: Markov Chain.
MCD: Markov Chain with Duration of Stay.
R: Renewal.

for the cohort sub-system. Table 6 provides measures for these comparisons. First, we use χ^2 for making *comparisons across grades per year*. The χ^2 statistic used here is viewed as a measure of agreement as in [16], rather than as a significance test since the flows between grades are interdependent (cf. [5]) and two different classes of models (Markov, renewal) with different inputs are being compared. The next to the last column shows the overall performance of the models for total system predictions in selected years. After both 5 and 10 years, the models may be ranked as follows: Renewal > Markov Chain > Markov Chain with Duration.

For *comparisons across time* of total system predictions, we use a simple average yearly relative error measure, shown at the bottom of Table 6. For the first five years, 1961–65, the MCD and R model are about equal at the COL grade, but for the other 4 grades the R model is more accurate. Surprisingly, the MC model is better than the MCD model at CPL and SGT grades and roughly equal at TPR and LT grades. It is only at the COL grade that the MC model fares poorly compared to the MCD model. Comparing the R and MC models, the R model's accuracy over this time period is consistently better, except at SGT where the MC model is one per cent better. For the total 10 year period, 1961–70, we observe from the last row of Table 6 that in general the Renewal model has the best fit and the simple Markov chain model appears slightly better than the Markov Chain with Duration.

As an overall measure across time and grade we will weight the yearly relative errors by size of grade and wage per grade. Size heavily weights the lower grades, whereas wage gives more weight to higher grades. More importantly, grade size times average wage per grade determines labor cost, a unitary and practical scalar for allocating errors across grades, thus providing an overall measurement across time and grades. The formula is simply

$$\frac{\sum_{i=1}^{k}\left[\left(\sum_{t=1}^{T}\frac{|P_i(t)-O_i(t)|}{P_i(t)}\right)\Big/T\right]\bar{O}_i W_i}{\sum_{i=1}^{k}\bar{O}_i W_i} \qquad (14)$$

where W_i denotes wage in grade i, $\bar{O}_i$ is the average grade size over the time period and $P_i(t)$ and $O_i(t)$ are the predicted and observed values of grade i in year t. Wages used were average grade level wage in 1970, the average across 5 seniority steps per grade. These salaries were COL: \$18,300, LT: \$14,300, SGT: \$12,200, CPL: \$11,000 and TPR: \$10,200. A one per cent weighted average relative error represents approximately \$150,000 per year for the 10 year period 1961–70 for the total system. The results are shown in the last two sets of rows and next to the last column of Table 6. For the five year period, there is no difference between the MC and MCD models. The renewal or vacancy model is clearly better than either the MC and MCD model. For the overall ten year forecasts, the models may be ranked as follows: R > MC > MCD with the MC model performing only slightly better than the MCD model.

For the cohort forecasts, the same three types of measures are shown in Table 6. For *comparison across grades* the χ^2 values in Table 6 show similar results as in the total system tests for 1961 and 1965. However, after 10 years the MCD model's overall accuracy is better than the MC model. The R model remains the best at both 5 and 10 years in the cohort tests. Similar findings hold for the comparisons across time, 1961–65 and 1961–70. In terms of the overall measure—the weighted average yearly relative error—the same results hold: over 5 years the MC model is better than the MCD model, but for the average 10 year forecast the MCD model is the better of the two. The R model remains the best of the three models for both 5 and 10 year periods.

For the second type of comparison, we compare the predictive errors between total system and cohort sub-system forecasts. Again, no statistical tests are conducted for the same reasons as before, as well as the fact that here the N's differ since the cohort is a subset of the total system. At issue is the range of predictions (aggregate, sub-aggregate) each model can be expected to handle and possibly directions for constructing a more adequate model.

First, note that in general the total system forecasts are more accurate than those of the cohort. This is as expected since the TS forecasts have more information in that recruits or growth are updated for the Markov and renewal models respectively. However, at specific grades the above finding (TS > CS) does not hold.

From the χ^2 values we note that the MC model becomes quite inaccurate for long term cohort forecasts, but is better than the MCD model for total system forecasts. The much better performance at the total system level further supports the observations made in Section 1 regarding the distinction between cohort and total system level testing for the MC model. The 10 year average yearly relative error data also support this observation, especially for the lower 2 grades where cohort depletions first occur. Finally, we note that the MCD seniority specifications improve the cohort forecast in the lower 2 grades, but that the MCD model fares poorly in the lower 3 grades against the MC model for the total system predictions. The cohort comparison suggests modest improvements by partitioning by seniority. More importantly, however, the total system level comparison suggests the same partitioning decreases predictive power. Specifically, the summation of cohort forecasts based on stationary duration mobility rates actually decreases accuracy in the 3 lower grades. The overall better performance of the R model at both levels is also potentially instructive. The improvement over either MC or MCD models at the cohort level suggests that perhaps seniority is operative in terms of ranking staff within grade, but that the number of persons moving from among this distribution is in response to available opportunities, thereby generating non-stationary duration mobility rates. In the R model, career movement is primarily viewed as dependent on the total system's vacancy multiplier process.

For the vacancy model, the 1961–1970 yearly relative error data indicate no difference between total system and cohort forecasting errors except at the bottom grade. Without lateral entry at the higher grades, almost all promotional vacancies arising in this 10 year period are filled by persons already in the system in 1960 (or the 1960 cohort). The larger cohort error in the bottom grade appears to be a function of the assumption of stationary loss rates and of the cumulative opportunity filtering property, whereby a vacancy at the top grade threads its way down to the TPR grade.

For the cohort predictions by the R model, the TPR grade is a residual staff not promoted and not exiting. Exit errors at any higher grade, however, eventually find their way to the TPR grade due to the hierarchical opportunity filtering process. Since the observed relative grade sizes in 1970 are sightly below those expected ($\hat{\rho}_j$) for all four upper grades, the job allocation error is in the opposite direction of the reported error (eg. in 1970 the R model overpredicted the number of TPR's remaining from the 1960 cohort). Hence, the exit error appears to account for most of the discrepancy. At issue here is that a Markovian cohort model's internal flows are in proportion to the *cohort's supply* whereas the renewal cohort model's internal flows are in response to the *total system's opportunities* or the demand side of the process.

5. CONCLUSION

To summarize, we have found the renewal or opportunity model to perform better in general than either the simple Markov chain model or the Markov chain model with duration of stay. This conclusion holds for both total system and cohort sub-system tests. As for each model's capability of handling the total system and cohort mobility behavior, the simple Markov chain model was inadequate for long term cohort forecasts, but much more adequate for the total system; both the Markov chain with duration of stay and the renewal models were better at the total system level than the cohort level; but each was much better than the simple Markov chain for long term cohort forecasts. The finding regarding the simple Markov chain model is consistent with earlier findings as reported in Section 1.

The overall better performance of the renewal type of model is particularly encouraging because of the crude input assumptions regarding exits and new job allocations. Thus, directions for even further improvement seem clear. Equally important, yet further theoretical refinements of the opportunity model *per se* have been developed (as yet unpublished). These advances include the incorporation of lateral recruitment at any grade, job abolishment, internal transfers within grades and managerial selection of individuals based on multi-attributes of candidates, with the choice conditional on the number of available opportunities. A promotional multiplier effect is also shown to be a property of the model, analytically deriving the number of promotions per entering vacancy. In sum, both the initial tests and additional theoretical developments seem particularly promising for the renewal/opportunity formulation.

Finally, we note the implications of the findings in this paper. First, these findings are consistent with those in [16] in which an alternative vacancy model also outperformed the Markovian type of model. Thus, on both empirical and theoretical grounds, we expect that most organizational staff flows are in response to the availability of opportunities or a pull mechanism. Unless staff movements are based on explicitly specified push mechanisms such as seniority, we suggest that much of the prior theorizing using simple Markov postulates [3, 5, 8, 12, 14 and 15] and work deriving therefrom [3] should now be reformulated to take explicity into account vacancies as in [1 and 16].

Vol. 6
1978

REFERENCES

1. Bartholomew DJ (1973) *Stochastic Models for Social Processes*. Wiley, New York.
2. Bartholomew DJ (1971) The statistical approach to manpower planning. *Statistician* **20**, 3–26.
3. Charnes A, Cooper WW & Niehaus RJ (1972) *Studies in Manpower Planning*. OCMM, Department of the Navy, Washington, DC.
4. Doeringer PB & Piore MJ (1971) *Internal Labor Markets and Manpower Analysis*. Heath, Lexington, Kentucky.
5. Forbes AF (1971) Markov chain models for manpower systems, in *Manpower and Management Science* (Eds. Bartholomew DJ & Smith AR). Heath, Lexington, Kentucky.
6. Henry NW, Mcginnis R & Tegtmeyer HW (1971) A finite model or mobility. *J. Math. Sociol.* **1**, 107–108.
7. Mahoney TA & Milkovich GT (1971) The internal labor market as a stochastic process, in *Manpower and Management Science* (Eds. Bartholomew DJ & Smith AR). Heath, Lexington, Kentucky.
8. Sales P (1971) The validity of the Markov chain model for a class of the civil service. *Statistician* **20**, 85–110.
9. Stewman S (1976a) Markov models of occupational mobility: theoretical development and empirical support, part 1: careers. *J. Math. Sociol.* **4**, 201–245.
10. Stewman S (1976b) Markov models of occupational mobility: theoretical development and empirical support, part 2: continuously operative job systems. *J. Math. Sociol.* **4**, 247–278.
11. Stewman S (1975a) An application of job vacancy chain model to a civil service internal labor market. *J. Math. Sociol.* **4**, 37–59.
12. Stewman S (1975b) Two Markov models of open system occupational mobility: underlying conceptualizations and empirical tests. *Am. Sociol. Rev.* **40**, 298–321.

13. WHITE HC (1970) *Chains of Opportunity*. Harvard University, Cambridge, Massachusetts.
14. YOUNG A (1965) Models for planning recruitment and promotion of staff. *Br. J. Ind. Rel.* **3**, 301–310.
15. YOUNG A & ALMOND G (1961) Predicting distributions of staff. *Comput. J.* **3**, 246–250.
16. YOUNG A & VASSILIOU RCG (1974) A non-linear model on the promotion of staff. *J. R. Statist. Soc.* **A137**, 584–595.

ADDRESS FOR CORRESPONDENCE: *Shelby Stewman Esq. Assistant Professor Sociology, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, Pennsylvania* 15213, *USA*.

1.6
78

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 353–361
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0801-0353$02.00/0

Graph Theory and Computer Aided Facilities Design

ALLAN S CARRIE
University of Strathclyde, Scotland

JAMES M MOORE
Virginia Polytechnic Institute and State University

MAREK ROCZNIAK
Krakow Polytechnic, Poland

JOUKO J SEPPÄNEN
Helsinki University of Technology

(*Received January 1978; in revised form March 1978*)

Four variations of an algorithm based upon graph theory are developed for solving the facilities design problem. The facilities design problem involves locating economic activities, each of which requires a given amount of space, on a floor within a building in such a manner to optimize an appropriate measure of effectiveness such as, for example, the total cost of materials handling. This problem has been called the location–allocation problem in some of the earlier management science literature. The algorithm uses strings to represent the graphs in order to be able to handle the computing more effectively. The algorithm is illustrated throughout the article by an example. Experience with both FORTRAN and PL/1 is reported. The topic is likely to be of interest to facilities designers, graph theoreticians and computer scientists.

1. INTRODUCTION

COMPUTERS have been applied to the problem of plant layout and facilities design for a decade. The availability of computers has led to the development of methods for quantifying the problem introducing new concepts far removed from the traditional visual charts. This use of the computer has introduced more objective measures of evaluation which have not been economic in the past. However, this application, like many computer-based techniques, may introduce a communication barrier between theory and practice, between the men who can obtain a solution and the men who must make it work. Probably the first hint that the mathematics of graphs might be useful in layout planning was by Levin [10]. A short time later, Krejcirik developed the RUGR Algorithm [6, 7]. In 1970 Seppänen and Moore [17] identified additional aspects of graph theory mathematics related to problems of facilities planning. Subsequent papers by the present authors, [2, 3], [12, 13], and [18], have further developed the application of graph theory to layout planning. This paper brings together these developments and presents the resulting procedure and algorithms.

2. GRAPH THEORY AND FACILITIES DESIGN

The facilities design problem referred to in this paper is the problem of locating a set of economic activities on a plane in such a manner as to optimize an appropriate measure of

effectiveness. The economic activities might be machine tools in a factory, desks in an office, departments in a service organization or workplaces of almost any type. The procedures described here are limited to locations on a plane or a single floor. The multiple story problem is not addressed in this paper although it is a non-trivial problem.

What is an appropriate measure of effectiveness? Only knowledge of the specific problem can answer this question. Some examples would be to minimize the total cost of materials handling in the factory situation, or to minimize the total cost of communications in the office situation, or to minimize the total flow of goods in the warehouse situation, or to minimize the total movement of personnel in the case of the service organization.

Although we chose to call this problem the facilities design problem, it has been called the plant layout problem by the Industrial Engineer. The economist and the management scientist often call this the location of economic activities. Undoubtedly other terms have been used to describe similar problems.

Let us use an example to present the procedure which has been developed. The subroutines on which the procedure is based will be described subsequently.

Step 1: Form a graph of the relationships between the activities in the problem

The activities involved in a facilities design problem can be thought of as the vertices in a graph. The relationships between them form the edges of the graph. The intensity of these relationships form the values of the edges of this relationship graph. The relationships are usually expressed in a chart or table, rather than diagrammatically, such as the Relationship Chart, due to Muther [14]. Figure 1 shows the data for an example problem in the form of a Relationship Chart. Following Muther's notation the relationship intensities have been graded into one of six categories and assigned code letters A,E,I,O,U or X. The value of each relationship is given at the point of intersection of the diagonals for the pair of activities concerned. For example, the relationship between activities A and C has been given an A category.

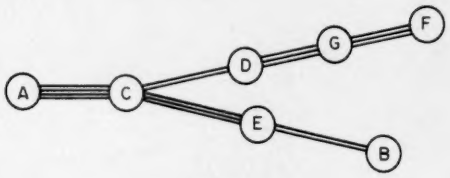


FIG. 2.

Step 2: Develop the maximal spanning tree of the relationships graph

A *maximal spanning tree* of a graph is the spanning tree the sum of whose edge values is a maximum. In facilities design, the maximal spanning tree includes the most important paths of flow. It contains the most essential relationships. Figure 2 depicts the maximal spanning tree for the example problem.

Relationship chart

	Activity	Area sq.meters
1	A	280
2	B	210
3	C	260
4	D	40
5	E	60
6	F	40
7	G	230

Plant brown Sc charted by JMM date 4-14

'Closeness' rating

Value	Closeness	Code in figures 2 and 4	Frequency
A	Absolutely necessary	|||||	1
E	Especially important	||||	3
I	Important	|||	3
O	Ordinary closeness OK	||	9
U	Unimportant		4
X	Undesirable	~~~	1
			21

FIG. 1.

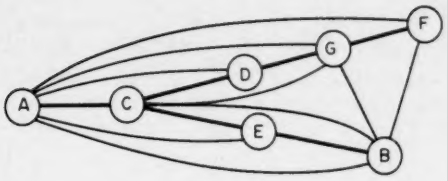


FIG. 3.

Step 3: *Develop a maximal planar subgraph of the relationship graph*

A planar graph is one which can be drawn on a plane surface so that no edges intersect. A *maximal planar graph* is a planar graph to which no other edges may be added without the graph becoming non-planar. It should be noted that such a subgraph is *not* necessarily unique. It should be further noted that a maximal planar subgraph is not necessarily the same as a maximum planar subgraph which is a planar subgraph with maximum total edge weight. The practical significance of the use of the maximal planar subgraph is pointed out later. Figure 3 shows the maximal planar graph for the problem. It has been impossible to include all the edges. Edge FC has been rejected.

Step 4: *Redraw the maximal planar graph to reflect the relationship intensities*

Figure 3 is only one of several ways in which the maximal planar graph may be drawn. It should be drawn so that the vertices sharing the most intense relationships are as close together as possible. The underlying theory and method used are described in the next section of the paper dealing with the development of the algorithms. Figure 4 shows the graph redrawn as an idealised Activity Relationship Diagram, using the codes defined in Fig. 1 for the relationship intensities, in the form used by layout planners [14].

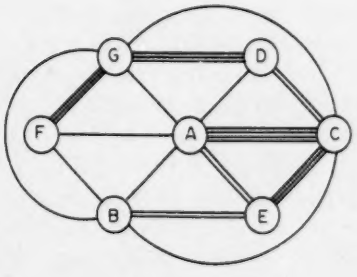


FIG. 4.

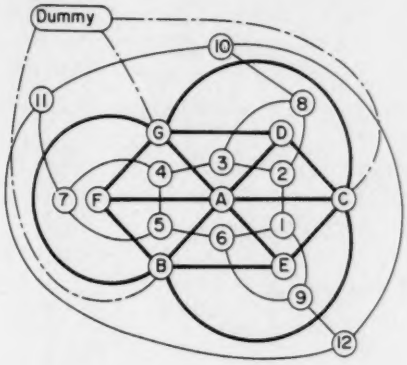


FIG. 5.

Step 5: *Draw the dual graph of the maximal planar graph*

The spaces between the edges of a graph are called *faces* and the *infinite face* is the area outside the graph. Every planar graph has a *dual graph*. If a point is placed on each face of the graph including the infinite face, then the dual graph has those points as vertices and has edges connecting those points whose faces were bounded by a common edge in the original graph. We introduce a dummy vertex which will become the area outside the layout in the end result, and which we connect to the activities on the periphery of Fig. 4. Figure 5 shows the dual graph in light lines. Figure 6 shows the dual graph with the maximal planar graph removed for clarity.

Step 6: *Form layout block plan*

Each face of the dual graph represents an activity in the layout. We need only "square-off", taking into account the area requirements of each activity to form the layout block plan, as shown in Fig. 7.

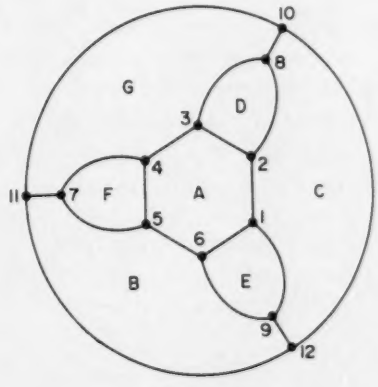


FIG. 6.

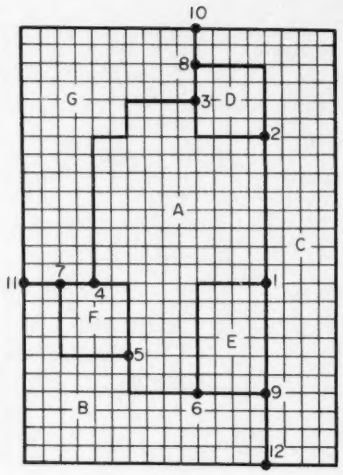


FIG. 7.

We have shown that a plant layout block plan is a planar graph, and that its dual is the graph of relationships between adjacent activities. The dual graph is also planar. It follows that if the complete graph of the relationships in the Relationship Chart is a planar graph, then there exists a block plan in which all the related activities are adjacent to each other. This block plan is a dual graph of the relationship graph. However, if the relationship graph is non-planar then there can be no layout in which all related activities are adjacent to one another. The reason for extracting a maximal planar subgraph from the relationship graph is now evident. The block plan which is its dual provides the maximum number of adjacencies between related activities.

3. DEVELOPMENT OF GRAPH THEORY ALGORITHMS

While graphs can be drawn on paper and analyzed with ease by human beings, it is only recently that efficient methods have been developed for processing them by computer. The maximal spanning tree is a well established concept and several algorithms have been published including Kruskal [8] and Seppänen [16]. Kruskal's algorithm for forming a maximal spanning tree is concisely stated and simple to follow manually. It says:

> "Among the edges which are not yet part of the tree, choose the edge which has the largest value and which does not form a closed loop with those already selected."

An alternative wording for seeking out the maximal spanning tree would be:

> Among the edges not yet in the tree, choose the edge which has the largest value and which connects a vertex not yet in the tree with one already in the tree.

Maximal planar graph

The concept of the maximal planar subgraph is not so well recognized as the maximal spanning tree. Testing the planarity of a graph is relatively well understood. Several algorithms have been proposed with those due to Tarjan [19] and Hopcroft and Tarjan [5] being the most efficient. To the best of our knowledge, no algorithm has been published on the extraction or construction of a maximal planar subgraph until this time.

One possibility is to amend or develop a planarity testing algorithm. However, a method based on Tarjan's algorithm, while giving generally good results, is too dependent on the nature of the particular problem. Nevertheless, we can formulate a theorem which will enable a maximal planar subgraph to be generated. The condition of the theorem is sufficient, but not necessary.

> Theorem: An edge can be added to a planar graph, so that the graph remains planar, if the chain bounding any face of the graph contains both vertices of the edge.

It follows from this that no edge can be added across a 'triangular' face consisting of three vertices and three edges. In a complete planar graph all faces are bounded by three edges, and are therefore triangular. Any face with more than three edges may be split up into smaller triangular faces by the addition of edges between non-adjacent vertices. Since each face has three edges, it follows that there is an upper bound (3V-6) on the number of edges which can be accommodated in a planar graph of V vertices. Consequently, the upper bound for the number of closeness relationships which can be satisfied by adjacency in a two dimensional layout is 3V-6.

Often there are more than 3V-6 edges in the relationship graph, but no more than 3V-6 of them can be included in any planar subgraph. In other words, all faces cannot be triangulated although edges remain, which cannot be added anywhere. Therefore, clearer distinction should

Vol. 6
1978

perhaps be made between two kinds of maximal (in terms of number of edges) planar subgraphs of the relationship graph:

(a) absolutely maximal (triangulated) planar subgraph, if such exists (this depends on the distribution of edges in the relationship graph), and

(b) maximal planar subgraph (not necessarily triangulated).

There may be several topologically different graphs of either type.

String representation

For processing of graphs by computer one needs some conventions for their representation in terms of linear strings of characters. A number of notations exist for this purpose. Perhaps the most common are the incidence matrix and adjacency matrix representations. However, for many specific kinds of problems, more suitable forms of representation can be devised. A special notation often offers great advantages, if it has closely analogous properties with the properties of the structures represented.

The crucial property in the facilities design problem is planarity. Seppänen and Moore [18] have introduced a string representation for trees and planar graphs, which has a one to one correspondence with these classes of graphs. Using this representation, plane embeddings of trees and planar graphs can be converted into a string language, which can be described in terms of very simple grammatical rules. Application of these grammatical rules allows one to generate layout topologies in terms of character string sentences. These can then be easily converted into graphical form. Perhaps the most important advantage is that the method allows one to totally circumvent the problems of planarity testing, because planarity is built into the grammar. In addition the method reduces manipulations of layout graphs into simple transformations of character strings.

The tree in Fig. 2 can be represented by the string

ACDGFGDCEBEC

which is produced by "walking around" the tree in a clockwise direction and noting the vertices encountered. By moving in the opposite direction, we would obtain

ACEBECDGFGDC

Notice that these two are topologically equivalent and that the string is unaffected by the vertex from which we start the process.

Manually, we add one edge at a time in order of decreasing weight and discard any edges which cannot be embedded. The addition of edges AE, AD, and BG would yield the graph of Fig. 8 whose faces are

ADC,
BECDG,
ACE and
AEBGFGD

Notice that edge FC cannot now be added without the graph becoming non-planar, because no face of the graph contains both vertices.

This figure may be used to illustrate another concept in the string representation, which will be utilised later in discussing the heuristic algorithms, that of the "distance" between two vertices in a face. This is the number of edges between the two vertices in the face string. For example, the distance between vertices D and B in face BECDG is three units measured via C and E, or two units measured via G.

Programming language considerations

Programming manipulations of character strings places special requirements on the programming language. Numerically oriented general purpose programming languages such as Algol 60 and FORTRAN are not convenient for dealing with structural problems and symbolic manipulations. Special purpose languages have been developed for such problems. The most notable of these is perhaps LISP [11]. LISP, short for List Processing, is a language for programming problems involving highly structured data and its manipulation. Data as well as the LISP programs themselves are

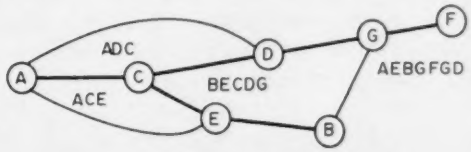


FIG. 8.

represented as parenthetic expressions of symbols called symbolic expressions. LISP would be very suitable for the present problem. However, the notation used in LISP programs, although elegant and universal, is perhaps a bit awkward even for experienced programmers.

Another well known language for symbolic manipulation is SNOBOL [4]. It is mainly designed for character string manipulation and natural language processing. SNOBOL would also do well for the present problem. More recently, however, general purpose languages have been developed, which allow both numerical computations and symbolic manipulations to be programmed in a conventional language framework. Such languages are PL/1, Simula 67, Algol 68 and POP-2. Of these, the two first named are more widely available. Results reported later on have been programmed in PL/1 to take advantage of its capability and in FORTRAN because of its universal availability even though it requires more programming effort.

Comparison of heuristics

An outline flow chart for the algorithm is presented in Fig. 9. It may be perceived from Fig. 8 that the sequence in which the edges are added to the graph can influence the number and value of edges which have to be excluded from the maximal planar subgraph. Edge FC had to be excluded because edges AD, AE and BG had been placed earlier. Had FC been considered before any of the others, FC would have been accepted and another edge rejected.

The way in which each edge is added is important. If edge BG had been added around the outside of F, the faces would be

BECDGFG and AEBGD

Instead of

BECDG and AEBGFGD.

It would still have been possible to add edge FC.

We therefore seek a heuristic rule for making these two decisions which will lead to good results. In fact, four decision rules have been developed as indicated in Table 1. Methods 1, 2, and 3 all add edges in descending order of flow value, but select positions in different ways. Seppänen and Moore [18] presented a worked example by Method 1 and pointed out that Method 2 would give better results. Edge BG is shown in Fig. 8 where it would have been placed by Method 2.

Method 3, developed by Roczniak, examines the possible locations and computes the apparent loss due to subsequent edges being excluded if the edge under consideration were to be added in each of the possible positions. Thus, Method 3 would have observed the apparent loss of FC if BC is added as shown in Figs 3 and 8, and the loss of AF if added in the alternative position discussed earlier. In this particular case, it would have been indifferent.

Method 4, developed by Carrie, is based on a different principle. With Methods 1, 2, and 3 all the faces so far generated must be scanned for possible locations. Method 4 seeks to add edges across the external face only. By adding edges, not necessarily in edge value order, to the external face, Method 4 tends to develop small, preferably triangular, internal faces which cannot be subdivided.

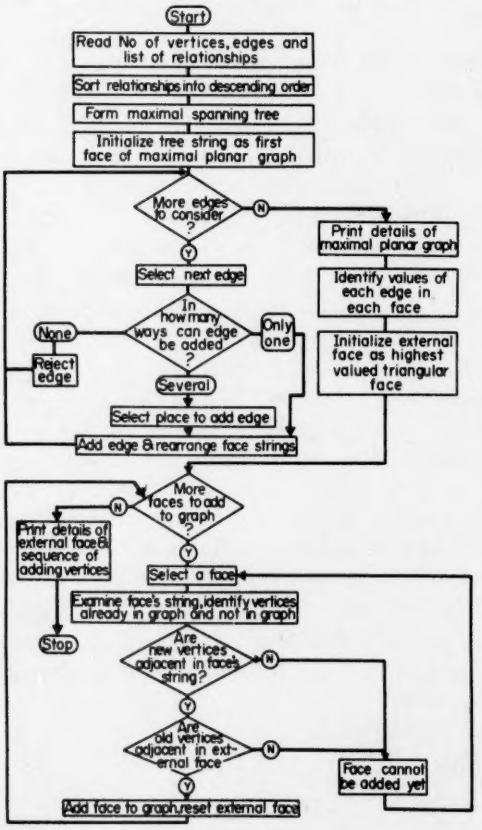

Start
Read No of vertices, edges and list of relationships
Sort relationships into descending order
Form maximal spanning tree
Initialize tree string as first face of maximal planar graph
More edges to consider ?
N
Y
Print details of maximal planar graph
Select next edge
Identify values of each edge in each face
In how many ways can edge be added ?
None
Only one
Several
Reject edge
Initialize external face as highest valued triangular face
Select place to add edge
Add edge & rearrange face strings
More faces to add to graph ?
Print details of external face & sequence of adding vertices
Select a face
Examine face's string, identify vertices already in graph and not in graph
Stop
Are new vertices adjacent in face's string?
Are old vertices adjacent in external face
Face cannot be added yet
Add face to graph, reset external face

FIG. 9.

TABLE 1. DECISION RULES USED IN FOUR HEURISTIC METHODS

Heuristic Method	Order of adding edges	Method of Selecting manner in which edge is added
Planar 1	Descending edge value	First possibility encountered
Planar 2	Descending edge value	Minimum Distance between vertices concerned, where distance = no of edges or vertices intervening
Planar 3	Descending edge value	Where apparent loss by exclusion of subsequent edges is minimised
Planar 4	1. Find minimum distance between vertices of any edge in external face. 2. Select highest value edge which can be added across this distance.	

Redrawing the maximal planar graph

The maximal planar graph defines the edges and faces to be used in planning the layout. A maximal planar graph may have up to 2v-4 faces including the external face. Two graphs which are topologically equivalent are said to be isomorphic. A complete planar graph can be drawn in 2v-4 different isomorphic ways, so that any face of the graph may be the external face. In the practical sense, this is significant because the external face does not form part of the layout. We therefore must draw the relationship graph in the most appropriate way for layout planning before forming the dual.

We do this by making the highest valued triangular face the nucleus of the layout and 'growing' the other faces onto it according to their edge value. The result is that the external face in the final layout is the most diffuse face in the graph. This part of the algorithm developed by Carrie is 196 FORTRAN statements, or 1136 words, in length.

4. EVALUATION OF USE OF GRAPH THEORY FOR FACILITIES DESIGN

Table 2 compares the algorithms according to various criteria and to solutions based upon Tarjan's planarity testing algorithm. We consider Method 3 to give very close to optimal values, however, Method 4 is faster and less dependent upon the initial maximal spanning tree string. The latter works better on problems whose trees have many branches.

Although algebraic compilors like FORTRAN can be used to process strings, they require a higher degree of effort on the part of the programmer than do the languages devoted to string processing. Table 3 makes a comparison between the use of FORTRAN and PL/1 for solving the problem of facilities design. Of course, the computer does more work with PL/1, but the demands on the programmer are reduced as a result. In FORTRAN, the programmer must use arrays to hold numerical equivalents of the strings. These arrays must be processed explicitly using DO loops.

Both the tree and the planar graph extract only a subset of the relationships which exist in a problem and their effectiveness will depend on the distribution of relationship values in the problem. Typically, not more than 20% to 25% of the relationships will be in the A or E categories and as many as 50% may be in the U category. Because of this distribution, the tree and planar graph can include a very high proportion of all important relationships.

Carrie [3] examined the distribution and effectiveness of the tree in five travel charts and found that on average 76% of entries were zero and only around 10% contained important values. The maximal spanning tree included on average 65% of all material handling and about 80% of the important material flows. The maximal planar graph was found to contain over 90% of all material handling and nearly 100% in some cases.

TABLE 2. COMPARISON OF MAXIMAL PLANAR GRAPH HEURISTICS

Heuristics		Planar 1		Planar 2		Planar 3		Planar 4		Tarjan 7	
Planar subroutine length											
No of Statement		97		119		150		250			
No of words		737		913		975		1985			
Core storage in 24 bit words (for up to 30 vertices)		9664		9856		10624		12480		18304	
	Number of										
Execution time in seconds	Vertices										
	7	0		1		1		0		3	
	10 (best)	0		0		1		0		16	
	10 (worst)	1		1		2		2		43	
	12	1		0		1		1		8	
	13	1		1		1		1		14	
	17	1		1		2		1			
	20	1		1		4		3			
	27	1		1		8		3			
Value of edges rejected, i.e. not included	7	2	93%	1	96%	91	96%	1	96%	1	96%
in maximal	10 (best)	302	88	218	91	178	93	227	91	300	99
planar graph	10 (worst)	489	81	337	87	239	91	286	89	454	82
and the % of	12	10	82	4	93	4	93	5	91	3	95
the total	13	218	91	24	99	17	99	17	99	19	99
flow value	17	1970	90	2085	90	2085	90	2045	90		
of graph	20	58	92	35	95	30	96	25	97		
included in	27 (best)					115	95	117	95		
MPG.	27 (worst)	192	91	184	91	143	94	138	94		

Note 1. The 10 department problem was run trying all of the 12 possible ways of describing the tree.

Note 2. The 27 department problem was run using three different strings.

Note 3. Comparison of program size and execution times for Tarjan 7 with the others is uninformative, because they refer to a complete program for 50 vertices and 100 edges, not planar subroutine only. Furthermore the runs were in the trace mode which increases run time by approximately a factor of ten.

A more recent study of 23 layout planning projects where the relationships were chosen by experienced engineers shows that, on the average, 62% of the ratings in these 23 projects were unimportant (U) ratings. The absolutely necessary or A ratings made up 5% of the closeness ratings on the average while especially important (E) ratings averaged about 6% of the ratings.

The upper bounds of feasible solutions are defined by graph theory which suggests that there is a clear cut goal when graph theory is the basis of the computer algorithm. Heuristics like CRAFT [1], CORELAP [9], and ALDEP [15] do not have the advantage of knowing the upper bounds. Carrie [3] compared the planar graph block plan with that produced by CRAFT for an example problem and found that the planar graph plan was superior with a layout cost of 89% of that of the CRAFT result. The definition of the upper bounds of feasible solutions may be the most important contribution to facilities design made by graph theory.

TABLE 3. COMPARISON OF FORTRAN TO PL/1

	Number of statements required	
	FORTRAN	PL/1
Planar 1	97	73
Planar 3	150	139

There have been objections to the heuristic approaches, because they have not allowed a user to examine the state of the design at the relationship diagram stage. Some feel that manual adjustment at this stage is useful and builds confidence on the part of the planners and decision makers. The graph theory algorithm terminates at the relationship diagram

stage permitting this manual adjustment. Steadman [20] has done some interesting work examining the conditions under which an effective layout can be realized considering the influence of room shapes.

REFERENCES

1. BUFFA ES, ARMOUR GC & VOLLMANN ET (1964) Allocating facilities with CRAFT. *Harv. Bus. Rev.* **42**(2), 136–159.
2. CARRIE AS (1973) Numerical taxonomy applied to group technology and plant layout. *Int. J. Prod. Res.* **11**(4), 399–416. Also in *Proc. of 2nd Int. Conf. on Prod. Res.*, Copenhagen 1973 (Eds. GUDNASSON & CORLETT). Taylor and Francis, London (1974).
3. CARRIE AS (1975) Graph theory applied to computer aided plant layout. *Proc. 15th Int. Conf. on Machine Tool Des. and Res.* (Birmingham). Macmillan, London.
4. GRISWOLD RE, POAGE JF & POLONSKY IP (1971) *The SNOBOL 4 Programming Language*. Prentice-Hall, Englewood Cliffs, New Jersey.
5. HOPCROFT J & TARJAN RE (1974) Efficient planarity testing. *J. ACM* **21**(4), 549–568.
6. KREJCIRIK M (1969) RUGR Algorithm, presented to the Computer Aided Plant Layout and Design Seminar, Technical University of Finland.
7. KREJCIRIK M (1969) Computer aided plant layout. *Comp. Aided Des.* **2**(1), 7–19.
8. KRUSKALL JB (1956) On the shortest spanning sub-tree of a graph and the travelling salesman problem. *Proc. of the Amer. Mathl Soc.*, pp. 48–50.
9. LEE RC & MOORE JM (1967) CORELAP-computerized relationship layout planning. *J. Ind. Enging* **18**(3), 195–200.
10. LEVIN PH (1964) Use of graphs to decide the optimum layout of buildings. *The Architects J.* Vol. 7.
11. MCCARTHY J (1962) *LISP 1.5 Programmers Manual*. MIT Press, Cambridge, Massachusetts.
12. MOORE JM (1976) Facilities design with graph theory and strings. *Omega* **4**(2), 193–203.
13. MOORE JM & CARRIE AS (1975) Impact of list processors and graph theory on use of computers for solving facilities design problems. *Proc. 3rd Int. Conf. Prod. Res.* Amherst. (Taylor and Francis, London, 1976).
14. MUTHER RA (1973) *Systematic Layout Planning* (2nd edn), Cahners, Boston Massachusetts.
15. SEEHOF JM & EVANS WO (1967) Automated layout design program. *J. Ind. Enging* **18**(12), 690–695.
16. SEPPANEN JJ (1970) Algorithm 399 spanning tree. *Communications of the ACM* **13**(10), 690–695.
17. SEPPANEN JJ & MOORE JM (1970) Facilities planning with graph theory. *Mgmt Sci.* **17**(4), B242–B253.
18. SEPPANEN JJ & MOORE JM (1975) String processing algorithms for plant layout problems. *Int. J. Prod. Res.* **13**(3), 239–254.
19. TARJAN RE (1971) An efficient planarity algorithm. *Report STAN-CS-244-71*, Stanford University, California.
20. STEADMAN P (1973) Graph theoretic representation of architectural arrangement. *Archit. Res. and Teaching* **2**(3), 161–172.

ADDRESS FOR CORRESPONDENCE: *Dr Allan S Carrie, Department of Production Management and Manufacturing Technology, University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, UK.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 363–366
© Pergamon Press Ltd 1978. Printed in Great Britain
0305-0483/78/0901-0363$02.00/0

Adapting Production for Private Label Goods

RS STAINTON

Henley, The Administrative Staff College, UK

(*Received January* 1978)

Linear Programming (LP) is a valuable tool of operational research and has been applied successfully to problems of production scheduling, particularly in the process industries. It is dependent to a large degree on forecasts of demand for the products: when demand is erratic, production scheduling suffers, whatever the technique employed. LP is sometimes adopted, however, in situations which do not necessarily warrant its use. Such a case arose when considering the production difficulties of private label goods in the food industry. The problem was eventually modified with the full cooperation of management, despite initial opposition. The solution found was to adapt as a company to market conditions, instead of forcing the production function to operate within artificial constraints.

l. 6
78

PRODUCTION PLANNING AND LP

DESPITE Bowman's [2] early reservations about using Linear Programming methods for production planning, the literature abounds with proposals for its use. He concluded his paper with the comment "though it is thought that this method has merit, it is quite difficult to get meaningful costs (linearity?), capacities (homogeneity?), and requirements (certainty?)".

The standard theoretical approach to the solution of production planning is now well documented, for example Nicholson and Pullen [7] and King [5]. Stainton [9] has made an attempt to extend the use of LP into realms of production scheduling as well as planning, but that method too relies upon various parameters which in practice might be difficult to quantify. Nevertheless, Jones and Rope [4] have described successful applications, as too have Williams and Redwood [10], among others.

Production planning in the process industries is probably more amenable to treatment by LP than it is in others. In a recent study undertaken with a large food manufacturer, there was evidence to show that LP was an appropriate method to apply to production planning. Unfortunately however, the situation was aggravated by the large range of private label goods to which the company had become committed. Mercer [6] has described some decision alternatives which might be pursued in order to consider whether to enter the private label market, but once the decision has been taken by a company to make private label goods for its supermarket customers, it is a far from easy matter to modify that strategy to any considerable degree.

PRIVATE LABEL AND SPARE CAPACITY

At one time, the company had been a major supplier of its own branded goods. In particular, it set up a modern and most advanced soup manufacturing plant in the full expectation that the market share for its products would increase and so fully utilize the extra capacity that it had created. In fact, due to competitive and other pressures, the reverse occurred and the company's share of the market began to diminish. In an attempt to correct the swing, the company agreed to supply one of its major supermarket customers with private label soups.

It was argued that the private label production would utilize marginal capacity and that, since the overheads would have already been recovered by production of the branded goods, the price charged to the customer should be

based on marginal costs as well. This meant that the soups could be sold by the supermarket at prices lower than those of the branded goods. In turn, it was hoped that this would enlarge the market for soups, or at worst eat into the sale of competitor products at the same rate as it would its own. Overall, the company should have gained an advantage.

At first the idea was successful, so much so that other supermarket customers wanted their own private label soups which the company was very willing to provide. The company's new plant began to be used almost to full capacity, although there was some concern that the sale of the branded goods were diminishing at a rate greater than had been expected. In an attempt to correct that situation, more effort was expended on advertising and promotion of the company's branded goods, with some effect. At the same time the company was pleased to take on more supermarket customers and to make private label soups for them.

SUPERMARKET BRAND RATIONALIZATION

A major difficulty arose when the supermarkets began to rationalize the number of products they kept on their shelves. The supermarkets reasoned that the loss of custom would be insignificant if the number of brands of each item that they stocked were reduced to three: fewer sales of one brand would be compensated by more sales of another with less complicated stocking procedures. As a result, it was decided to continue to hold the leading brand, the next most popular or the one which was currently increasing in sales, and their own private label.

In the great majority of cases this meant that the food manufacturing company's own brand of soups was excluded from the major retail outlets and it became dependent immediately upon the success of the private label goods which it manufactured. The supermarkets argued that since the private label soups were identical to the branded ones, the company was losing very little, if anything. Equally, since the company would not need to advertise so intensely in future and therefore overheads would be correspondingly reduced, there was no case for increasing prices for the private label soups above those already agreed on the marginal costing basis. Besides, if the company were to refuse to supply the private label soups, it could not make up the deficit by selling its own branded goods through those same outlets; neither would the supermarkets be inclined to continue to purchase many of the other goods which the company manufactured and distributed.

PRIVATE LABEL PRODUCTION

The situation had arisen therefore that by far the greater proportion of the company's soup production was for private label, with a very small demand for its own branded goods. The prices charged by the company for the private label soups were too low, but there was considerable resistance to any proposed increases.

The manufacturing plant had been designed with just one brand in mind and was therefore a continuous process of canning, including labelling. At first sight, this would seem to suit well the production of private label goods since the recipes for each supermarket are the same. But the labels are unique to each supermarket and in consequence, it was necessary to plan production by label as well as by soup type: one supermarket's chicken soup, for example, became quite a different product to another supermarket's chicken soup even though the contents of the cans of each were identical.

In attempts to increase the sales of canned soups, the marketing division had entered into a large variety of restrictive contracts. They had agreed, for example, to fixed prices for fixed terms, to deliver the goods to retail outlets or to supermarket central depots as the customer wished, and to supply whatever quantities the supermarkets required on demand.

It soon became apparent that with the proliferation of private label soups, the contractual obligations of supply could not be met. It had become exceedingly difficult to plan production to satisfy demand, even more difficult to do so at minimum cost. Nevertheless, a production planning analysis was begun and it was not long before the linear programming tools were introduced. Immediately, all of the reservations raised by Bowman [2] were

encountered and efforts were made to overcome them. The costings developed by the accountants were embarrassing and became a political issue: as a result, those finally presented were suspect. Plant capacities were the least difficult to handle, but forecasting the uncertainties of demand proved to be a considerable problem.

A team was set up to analyse past and potential demand data in detail and various forecasting techniques were explored. Eventually, a method was agreed and adopted for an LP formulation of the production planning process. The results obtained purported to identify when each private label soup should be manufactured, the quantities and at minimum cost. Regrettably, service to the customers was little better than it had been previously, various crisis action had to be taken at shop floor level and hence costs were still relatively high. More effort was devoted to improving the forecasts.

REDEFINING THE PROBLEM

Rivett [8] discusses how the order in which evidence is presented to the problem solver influences his understanding of the problem and the manner in which he might tackle it. He perceives the situation in a particular way and employs the tools at his disposal. Another problem solver might recognize the problem in quite a different way and arrive at a different solution.

If we were to stand back and look at the production planning of private label soups more clinically, we might recognize that contrary to the company's protestations and wishful thinking, it is no longer in the business of making branded soups, but private label ones instead. However painful it might be, it must adapt to the reality of that situation. Ackoff [1] has introduced the concept of adaptivising, although Eilon [3] argues that it is no more than satisficing. Either way, it is a problem solving process which could bear fruit in the situation which has been described.

If the problem solver (operational researcher) is to be truly effective he must recognize the proper constraints of his problem and the limitations. Eventually, it was realized and accepted that within the terms of trading between the company and its supermarket customers, suitably accurate forecasting could not be achieved, thus creating considerable difficulties for production planning based on those forecasts. The operational research team working on the problem took a courageous decision to lay aside their tools of trade, thereby exposing themselves to some managerial scepticism, and to approach management instead with the suggestion to rethink the problem.

Two proposals were placed before management, although others, such as to pull out of soup production entirely, might equally well have been discussed. The two chosen were considered to be realistic, provided management were prepared to re-orientate to the private label soup production business. The proposals were:

1. to redesign and therefore reconstruct the manufacturing plant so that shiners (filled cans without labels) be produced and subsequently stored for future off take. This would alleviate the forecasting problem since it was identifying the demand of each individual private label which caused the greatest difficulty: forecasting by soup type was less of a problem. When required, the shiners would be labelled by supermarket, at an additional cost of handling out of store and onto new labelling plant;

2. to negotiate new contracts with the supermarket customers so that they would take fixed quantities of stock at regular intervals. In order to persuade the customers to adopt these new terms, the company would need to offer a more attractive discount structure.

Neither of these proposals appealed to the company's management, but eventually the reality of the situation was grudgingly accepted. Although both the proposals would result in losses for the company, it was argued that those losses would be less than those currently incurred. Since management was not prepared to pull out of soup production in its entirety, for a variety of related reasons, it was necessary for it to operate simply at an acceptable level of loss.

NEW TERMS OF REFERENCE

The operational research team were instructed to investigate in greater depth the

alternatives it had proposed, and to co-operate with management in a search for others which might even be preferable. Instead of being a service function for management which had been constrained to operate within the artificial boundaries of an initial problem definition, they had broken out to identify with management what the problem really was, and to take steps to convince the management that a new approach to solution was needed.

It is tempting to stay within well trodden paths of LP, forecasting and other methods, but it is not necessarily in the interests of the company as a whole to do so. Furthermore, management itself may well take some persuading to look to other ways of solving its problems: it may prefer to pass on the difficulties to another department and feel comforted when that department is unable to solve the problem with its sophisticated techniques. To tread new ground by recognizing a problem as something other than that which was readily defined, and to recognize that its solution requires radical changes to be made, is sometimes less than appealing to a management concerned primarily with short-term success. Therein lies a dilemma for operational research: to battle on with the techniques or to adapt not only its own thinking but that of the management whom it serves.

REFERENCES

1. ACKOFF RL (1970) *A Concept of Corporate Planning.* John Wiley, New York.
2. BOWMAN EH (1956) Production scheduling using transportation method. *Op. Res.* **4**, 100–105.
3. EILON S (1972) Goals and constraints in decision making. *Opl Res. Q.* **23**, 3–15.
4. JONES WG and ROPE CM (1964) Linear programming applied to production planning—a case study. *Opl Res. Q.* **15**, 293–302.
5. KING JR (1975) *Production Planning and Control.* Pergamon Press, Oxford.
6. MERCER A (1966) Applications of operational research in marketing. *Opl Res. Q.* **17**, 235–252.
7. NICHOLSON TAJ and PULLEN RD (1974) *Computers in Production Management Decisions.* Pitman, London.
8. RIVETT BHP (1972) *Principles of Model Building.* John Wiley, London.
9. STAINTON RS (1977) Production scheduling with multiple criteria objectives. *Opl Res. Q.* **28**, 285–292.
10. WILLIAMS HP and REDWOOD AC (1974) A structured linear programming model in the food industry. *Opl Res. Q.* **25**, 517–528.

ADDRESS FOR CORRESPONDENCE: *Dr RS Stainton, Henley, The Administrative Staff College, Greenlands, Henley-on-Thames, Oxfordshire RG9 3AU, UK.*

MEMORANDA

1. 6
978

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 4, pp. 369–378
© Pergamon Press Ltd 1978. Printed in Great Britain

0305-0483/78/0901-0367$02.00/0

Travelling Salesman Distances

IN MOST practical vehicle scheduling or travelling salesman problems the use of euclidean distances does not adequately reflect the true distances travelled even when a multiplicative factor is used. This fact is recognised by the emphasis placed on the use of a barrier and passpoint system, or more recently a road network in such problems. Even then it is not obvious which distances or routes should be used to develop the shortest tour. Consider the symmetrical distance matrix shown in Table 1.

TABLE 1.

from \ to	A	B
B	2	
C	7	3

The travelling salesman tour round the three points using the above matrix is A–B–C–A, a distance of 12 units. However, the shortest route is A–B–C–B–A, a distance of 10 units. Clearly the second route violates the constraint that each point is visited once and once only, and hence would be difficult to determine using normal travelling salesman techniques. This problem can, however, be solved by using a shortest route algorithm, for example that described by Floyd [1] to determine the distance matrix prior to solving the travelling salesman problem. Using the approach on the example, gives the distance matrix shown in Table 2, from which the shortest route is readily attainable.

TABLE 2.

from \ to	A	B
B	2	
C	5	3

This method was also applied to the 10 symmetrical 9 city problems of Robacker [2]. Comparative results are given in Table 3 below. Here a travelling salesman routine (TSRP) due to Spears [3] (this is a branch and bound algorithm similar to that of Little *et al.* [2], but using depth first, and throw away tree search) was employed on Robacker's original matrices and then modified by Floyd's algorithm.

It can be seen from Table 3 that the use of the approach described above can have considerable impact on the value of the final solution.

REFERENCES

1. FLOYD RW (1962) Algorithm 97 shortest path. *Comm. of ACM* **5**, 345.
2. LITTLE JDC, MURTY KG, SWEENEY DW and KAREL C (1963) An algorithm for the travelling salesman problem. *Op. Res.* **11**(6), 972–989.
3. ROBACKER JT (1955) Some experiments on the travelling salesman problem. *Rand Corporation Research Memo*, **RM-1521.**
4. SPEARS R (1968) The Travelling Salesman Problem. Unpublished M.Sc. Thesis, Department of Management Science, Imperial College, London.

Prof. KB Haley
CDT Watson-Gandy

Imperial College
Department of Management Science
Exhibition Road
London SW7 2BX
UK

(*December* 1977)

TABLE 3. RESULTS OF 10 TRAVELLING SALESMAN PROBLEMS† WITH AND WITHOUT USE OF THE SHORTEST PATH ALGORITHM.

Problem No.	Robacker's solution	TSRP solution without SPA	TSRP solution with SPA	% Improvement by SPA
1	232	232	224	3.7
2	204	204	171	19.3
3	178	178	137	29.9
4	185	181	168	7.7
5	283	283	274	3.3
6	150	150	145	3.5
7	185	185	161	14.9
8	159	159	128	24.2
9	266	266	262	1.5
10	236	236	164	43.9

† Problems taken from [2]
TSRP = Travelling salesman routine
SPA = Shortest path algorithm

Management Science in Developing Countries: Adopting the Right Style

MOST PEOPLE whose views I have heard or read recently agree that MS could have a big impact on development. But there is much less agreement about the style of MS which should be adopted: particularly, priorities in terms of methods to be used and where to apply them. Here, I briefly describe three styles which have often been proposed, and criticize them based on my six years as an MS man in developing countries.

The first recommends concentration on the use of established analytical techniques. Proponents claim that there are many easily identifiable applications in developing countries, and quick results can be obtained of demonstrable success.

The second stresses the need to get large numbers of people using simple techniques. Michael Luck suggests the term 'barefoot MS'. It is assumed that the main difficulty lies in identifying and formulating the problems, which can then be solved without need for sophisticated analysis.

The third is based on the claim that problems and their solutions are well known, but implementation of change is constrained by group attitudes and traditions. Russel Ackoff [1] says that the first job is to work on cultural constraints such as paternalism, corruption, and conformity. Until the system is amenable to change, there is no useful role for analysis of alternatives.

I am convinced that each of these styles is important, and selection (or combination) depends on the details of the problem and its environment. For example, it may be that use of an established analytical technique is valuable since it helps to establish the soundness of MS as a whole. Similarly, it is often important to ensure that large numbers of persons understand the basic elements of MS (and have done some simple analyses or their own), before attempting any revolutionary change affecting self-imposed constraints to development. And it cannot be doubted that any small scale work, or that involving use of standard mathematical models must be done with prior understanding of the best way to achieve rapid development.

Our strategy for training and application in the Indonesian health services involves all three styles. For example, we now have a small team of MS specialists applying standard techniques such as linear programming to a wide variety of problems (for example, a caterer formulation for drug production: and a classical diet formulation for nutrition planning and education). At another level, we are halfway through a programme which will give every manager from health centre doctor upwards a three week course in MS, followed by supervision of projects in their normal places of work. Finally, we are attacking the cultural constraints at various levels. For example, we have introduced Delphi methods to improve decision making. The traditional method of allowing the senior person complete power is being replaced by participatory planning and policy formulation involving persons at all levels.

In summary, these styles are interdependent. It is the nature of the discipline that they should be so. One last example illustrates the point. A developing country wishes to determine an appropriate balance of resource utilisation between hospital and health centre services. An LP formulation can be used, where there is a matrix of potential clients and service types, and the cost elements are group judgments on expected value derived by person i from service j. This kind of project requires that many people understand basic concepts, so they can participate in refinement of the formulation and provision of data, and will want to help implement. But it also requires that we recognise the cultural constraints. For example, most developing countries spend 80% or more of their resource on hospitals. This is partly because capital expenditure gives the opportunity for corruption—a percentage. At health centre level and below, the opportunities are rare. For example, the major resource is voluntary activities by the villagers themselves.

This kind of problem must be attacked from all sides. The MS scientist must, as Samuel Eilon [2], puts it "be concerned with the main issues", but he will find it of great advantage to be technically competent, and to be working with people who have a basic understanding of scientific method.

REFERENCES

1. ACKOFF RL (1977) National development planning revisited. *Op. Res.* **25**, 207–218.
2. EILON S (1977) Technician or Adviser? *Omega* **5**(1), 1–6.

Dr Don Hindle

(*January* 1978)

Health Services Research Institute
Jalan Indrapura 17
Surabaya
Indonesia

Vol. 6
1978

Estimation of a Dirichlet Prior Distribution

BACKGROUND

THE USE of the beta prior distribution in Bayesian analyses involving the Bernoulli parameter is now becoming quite common. It is the natural conjugate prior distribution for a Bernoulli, or binomial, data process (e.g. Lindley [5] gives a good exposition of this) and is therefore very useful whenever prior subjective belief is to be articulated on the occurrence of one of two dichotomous events. In a marketing context, these dichotomous events could be the purchase or nonpurchase of a particular product by an individual. Thus Murray and Silver [6], for example, have found the beta distribution useful in market share forcasting. In a financial context, Bierman and Hausman [1] utilised a beta distribution for a credit granting decision to model the payment of debts.

The multivariate version of the beta distribution is the Dirichlet, which is thus the natural conjugate in the Bayesian analysis of the multinomial data process. Therefore, if the sample space consists of n mutually exclusive and collectively exhaustive events, e_i, then the Dirichlet density function can represent subjective belief, via the simplex of probabilities, p_i, on the probability of event e_i.

$$f(p_1, \ldots, p_n) = \frac{\Gamma(\Sigma a_i)\Pi(p_i^{a_i - 1})}{\Pi(\Gamma a_i)} \tag{1}$$

with $\Sigma p_i = 1$ and all a_i positive.

The posterior mean of p_i, after j sampling is

$$\bar{p}_i = (a_i + s_{ij})/(j + \Sigma a_i) \tag{2}$$

where s_{ij} denotes the number of times out of j that event i occurred.

The Dirichlet formulation has not been widely used in the practice of decision analysis. This may be because of the apparent mathematical complexity involved, although this can be considerably alleviated with each p_i being marginally distributed beta, $B(p_i | a_i, \bar{a}_i)$ where

$$a_i = \sum_{j \neq i} a_j \tag{3}$$

although the constraint $p_i = 1$ imposes only $n - 1$ such independent marginal distributions.

ESTIMATION

A major difficulty in using a Dirichlet prior distribution is the practical problem of its subjective assessment. Unlike the beta distribution, it is not univariate and therefore is not amenable to the usual fractile assessment procedures, etc. (e.g. Bunn [3] for details). The parameters could be obtained by assessing the marginal beta distributions, but then difficulties may be encountered in reconciling the n parameters, given the constraint of equation (3).

An alternative method is one of 'imaginary results' suggested by Good [4]. There are two versions of this methodology; one known as Equivalent Prior Samples (EPS), the other is the Hypothetical Future Sample (HFS). EPS is simply the number of samplings that the assessor thinks would be the empirical equivalent of his subjective belief. Thus, he would give a set of prior means $\bar{p}_i^*$ and a number v which represents the amount of data that he thinks would be equivalent to his subjective state of information. Hence, the prior parameters are obtained as

$$a_i = \bar{p}_i^*(v + n) \tag{4}$$

Incidentally the multiplicative factor is $(v + n)$ and not v because an initial uniform uninformative prior density function is envisaged (with all $a_i = 1$) for reasons analagous to those of Bunn [2] relating to the uninformative beta distribution.

The methods of EPS would appear particularly vulnerable to the cognitive assessment biases discussed by Tversky [8] because of the conceptual demands it makes on the assessor. HFS appears at first slightly more tolerable in this respect. The assessor gives prior expectations, $\bar{p}_i^*$, as before, but is then asked what his posterior expectation $\bar{p}_i^{**}$ would be if the event i occurred k times in the next m samplings ($m \geq k \geq 1$).

Now from equations (2) and (4)

$$a_i = \bar{p}_i^*(k - m\bar{p}_i^{**})/(\bar{p}_i^{**} - \bar{p}_i^*) \tag{5}$$

HFS clearly makes very strong assumptions on the assessor being an intuitive Bayesian and the method may well be subject to assessment bias in this respect. Research on 'conservatism' in human information processing, by for example Phillips *et al.* [7], has indicated that individuals tend to revise their probabilistic beliefs insufficiently with the impact of new data compared with Bayes theorem. More recent work on 'conservatism', however, has not been unequivocal and there is the general feeling (see Winkler [10] or Youssef and Peterson [11] for example) that individuals are not so suboptimal in their real world activities as in the more hypothetical, experimental conditions of the psychological laboratory. In a comparative evaluation for assessing a beta distribution, Winkler [9] did find that the assessment bias in the imaginary results methods was slightly greater than in the usual fractile method, although the subjects did express a personal preference for the methods of imaginary results. Again, the hypothetical nature of experimental research in this area militates against its generalisability to real world decision making. In a case described below, a method mathematically similar to HFS was used to assess a Dirichlet distribution except that actual, rather than hypothetical, sample results were used. This method could be called Actual Future Sample (AFS) and, whilst still subject to possible conservatism bias, should not be affected by the hypothetical contrivance of the experimental laboratory. The presentation of this case indicates the feasibility of the method and some elementary insight into its possible assessment bias properties.

CASE STUDY

Towards the end of 1977, a Los Angeles publisher was offered a diversification opportunity to launch a new product through his existing consumer outlets. For his decision analysis, he desired to partition the outcome space into three events. If x is the number of his consumer outlets who will take the product, then

e_1: $x \leq 10\%$
e_2: $10\% < x \leq 25\%$
e_3: $x > 25\%$

He gave a set of prior expectations for each of these events, viz.

$$p_1^* = 0.20$$
$$p_2^* = 0.30$$
$$p_3^* = 0.50$$

The EPS was assessed as $v = 7$.

Twenty of these customers were then interviewed at random. Six of them believed that the product would capture less than 10% of the market, i.e. e_i; 7 for e_2 and likewise 7 for e_3. These were treated as samplings from each of the three categories (the total number of customer outlets was about 300). The publisher was then asked what his revised expected probabilities for the three events were now evaluated as on the basis of this new information alone. The following values were assessed.

$$\begin{aligned} \bar{p}_1^{**}(A) &= 0.25 \\ \bar{p}_2^{**}(A) &= 0.30 \\ \bar{p}_3^{**}(A) &= 0.45 \end{aligned} \tag{7}$$

where the parentheses indicate that these posterior probabilities were derived by subjective assessment (A) rather than computation (C).

Now, if the assessor had been a perfect Bayesian, the posterior probabilities should have been those computed via Equation (2), viz

$$\begin{aligned} \bar{p}_1^{**}(C) &= 0.27 \\ \bar{p}_2^{**}(C) &= 0.33 \\ \bar{p}_3^{**}(C) &= 0.40 \\ v &= 25 \end{aligned} \tag{8}$$

The differences between $\bar{p}_i^{**}(C)$ and $\bar{p}_i^{**}(A)$ indicate that the extent of the assessor's probability revision was insufficient, i.e. a 'conservatism' phenomenon. These results could also be explained through some basic EPS assessment bias in estimating too small a v.

An alternative analysis can be undertaken using the AFS idea. Using equation 5 on $\bar{p}_1^*$ and $\bar{p}_1^{**}(A)$, we have

$$\begin{aligned} a_1 &= 0.20(6 - 5)/(0.25 - 0.20) \\ &= 4 \end{aligned}$$

and thus from equation (4),

$$a_1 = 4;\ a_2 = 6;\ a_3 = 10 \quad \text{i.e. } v = 17. \tag{9}$$

Using equation (5) on $\bar{p}_3^*$ and $\bar{p}_3^{**}(A)$, we have

$$\begin{aligned} a_3 &= 0.5(7 - 9)/(0.45 - 0.50) \\ &= 20 \end{aligned}$$

and thus for equation (4),

$$a_1 = 20;\ a_2 = 30;\ a_3 = 50 \quad \text{i.e. } v = 97. \tag{10}$$

The disparity in the results of equations (9) and (10) reflects the inconsistency in the assessed posterior distribution. The assessor revised the posterior probabilities insufficiently in the light of the new data. This finding support the conservatism hypothesis but it is important to realise the level of accuracy in measuring the prior and posterior expectations that formula 5 imposes. Recall that

$$a_i = p_i^*(k - m\bar{p}_i^{**})/(\bar{p}_i^{**} - \bar{p}_i^*) \tag{5}$$

It is clear that measurement error (as opposed to assessment bias) in computing the expectations in the denominator, particularly if $(\bar{p}_i^{**} - \bar{p}^*)$ is small, can cause large errors in the derived a_i.

Furthermore, if k exactly equals the posterior expectation $m\bar{p}_i^{**}$, then the numerator becomes zero. Additionally, since a_i is constrained to be positive, then whenever $k = m\bar{p}_i^{**}$, $\bar{p}_i^{**} = \bar{p}_i^*$ and a_i is undefined. This result is consistent with coherent reasoning. If the actual sample is equal to the posterior expected sample, then the prior and posterior expectations are the same and we can learn nothing about the strength (i.e. information content, v) of the prior beliefs. The sample result was unsurprising and caused no change in expectations.

The converse of this should also hold, if the assessor sees fit not to change his prior expectation, then the sample should have been exactly as expected.

Again, the assessor in this case-study was inconsistent in this respect for keeping $\bar{p}_2^*$ unchanged at 0.3 despite having 7 rather than 6 'events' in this category.

SUMMARY AND COMMENTS

It is unrealistic to expect a probability assessor to be perfectly consistent, particularly under a method so sensitive to measurement error as AFS. However, the extent of the disparity in the AFS results and also with the EPS result in the above case study together with the evidence of conservatism requires some further comment.

Conservatism in intuitive Bayesian reasoning can also be explained if the assessor believes the data to be less than perfectly reliable. Clearly if the assessor believes the new data to be degraded in some way, he will not revise his prior probabilities by as much as if it were believed to be perfectly diagnostic. In the above case, the assessor thought that the opinions expressed in the sample may have been biased by the personal perferences of the individuals being sampled. Such unreliability of intentions surveys is a common problem in marketing research and a difficult one to model explicitly. Thus, in the above case-study, whilst the decision-maker found the Bayesian analysis useful in clarifying his reasoning, he still chose to stay with his assessed posterior expectations because of factors not explicitly considered in the AFS and EPS methods.

In summary, because of the mulitvariate nature of the Dirichlet prior distribution, the application of the usual univariate fractile methods of estimating the parameters may generally be difficult and tedious in practice. Other methodological candidates include assessing the equivalent prior sample (EPS) directly or imputing it from the way in which expectations are revised according to new sample information, either hypothetical (HFS) or actual (AFS). These latter methods may require several iterations of review on the part of the assessor to achieve consistent results.

REFERENCES

1. Bierman H and Hausman WH (1970) The credit granting decision. *Mgmt Sci.* **16**(8), B519–B532.
2. Bunn DW (1976) Bayesian point estimation on the Bernoulli parameter. *Omega* **4**(1), 97–99.
3. Bunn DW 1978) *The Synthesis of Forecasting Models in Decision Analysis.* Birkhauser Verlag, Basel, Switzerland.
4. Good JJ (1965) *The Estimation of Probabilities.* MIT Press, Cambridge, MA.
5. Lindley DV (1965) *Probability and Statistics.* Cambridge University Press, Cambridge.
6. Murray GR and Silver EA (1966) A Bayesian analysis of the style goods inventory problem. *Mgmt Sci.* **12**, 785–797.

Vol. 6
1978

7. PHILLIPS LD, HAYS WC and EDWARD W (1966) Conservatism in complex probabilistic inference. *IEEE Trans. on Human Factors* HFE-7.
8. TVERSKY A (1974) Assessing uncertainty. *J. R. Statist. Soc. B* **36**, 148–159.
9. WINKLER RC (1967) The assessment of prior distributions in Bayesian analysis. *J. Am. Statist. Soc.* **62**, 776–800.
10. WINKLER RC and MURPHY AH (1973) On the generalizability of experimental results. *IIASA Research Report* RR-73-15, Austria.
11. YOUSSEF Z and PETERSON C (1974) Intuitive cascaded inference. *Organisational Beh. and Hum. Perf.* **10**(3), 349–358.

Derek W Bunn

(*February* 1978)

School of Business Administration
Department of Quantitative Business Analysis
University of Southern California
University Park
Los Angeles
CA 90007
USA

1. 6
78

An Implicit Enumeration Algorithm for Solving Zero-One Goal Programming Problems

IN AN EARLIER paper [2] a description was given of how the production of improved student nurse allocation programmes could be formulated as a goal programming problem involving zero-one variables. The need then arose for an algorithm capable of solving large scale (e.g. 200 × 300) zero-one goal programming problems.

Due to the nature of the student nurse allocation problem it was anticipated that the priority structure of the goals could, and in all probability would, change in a dynamic fashion as the search routine developed. Consequently, the approach adopted was one which investigated the Decision Maker's implicit utility function rather than one which relied upon an *a priori* decision about its fundamental composition (e.g. see [4] or [6]). The method involves the use of a weighting vector **w** to represent the Decision Maker's utility function at any particular time.

The algorithm devised for solving the problem once **w** has been established is an iterative search procedure which implicitly evaluates all possible solutions. As only extreme points of the unit hypercube are considered termination in a finite number of iterations is ensured. The algorithm performs efficiently primarily because advantage is taken of the zero-one characteristic of the problem which serves to considerably simplify much of the computational work involved. Indeed, it is fair to describe the mechanics of the method as being of a relatively trivial nature and certainly at each stage far less complex than those involved in a Simplex iteration of a corresponding magnitude. Each goal is defined and an implicit objective function of minimizing the sum of the weighted deviations from each attainment level set up. At each iteration the algorithm determines which goal has the largest weighted deviation from its desired level. Termination will occur if the value of the objective function is zero, as this indicates that all goals have been met exactly.

The goal selected is scanned to identify the variable whose introduction to, or removal from, the current basis will most nearly achieve that goal's desired attainment level. To ensure that the effect on the overall objective function of introducing (removing) the variable into the basis is definitely not unfavourable a value similar in interpretation to that of the reduced cost figure in the Simplex method is calculated. This pseudo-cost figure for each variable is calculated by premultiplying its column vector of goal coefficients by the row vector of goal weights. We note that although the calculation of this figure (hereafter referred to as Z value) is identical in process to the calculation of the reduced cost figure in a Simplex iteration it does not carry the same precise message. Due to the oscillatory nature of this algorithm about all the goals a negative Z value does not necessarily guarantee an improvement in the objective function as 'moving towards a goal' also includes travelling beyond it. However, the introduction of a variable having an associated non-negative Z value will definitely produce an inferior solution. The use of the Z figure does not exclude combinations of variables entering the basis to improve the sum of weighted discrepancies. If a combination of variables would effect such an improvement then at least one of the individual variables will have a negative Z value, thus allowing it to enter the basis. If one member of an attractive combination has been introduced then the others will follow, by the same argument.

Once the incoming (outgoing) variable has been identified an updating operation must take place. Due to the strictly 0/1 nature of the algorithm and the implicit consideration of the slack or discrepancy variables the only vectors which change at any one iteration are those corresponding to the incoming (outgoing) variable and the discrepancy vector. Obviously the discrepancy vector will be reduced by the vector of the incoming (outgoing) variable and the latter will now have to represent the effect of that particular variable when leaving (entering) the basis and thus must become the negative of itself. It may be that after a particular iteration one or more of the right hand sides have become negative. This is perfectly acceptable as it simply represents an over rather than an under achievement of any particular goal. But for computational purposes the row corresponding to any goal which has a negative right hand side is multiplied through by -1 to re-establish a positive right hand side. In cases where under or over achievement of goals are not viewed with equal severity, minor changes in the algorithm are necessary and will be dealt with later.

Due to the myopic nature of the solution technique whereby only one goal is considered at each iteration the likelihood of cycling is quite high. In order to prevent and also capitalise on this, it is necessary to compare at all stages the existing basis with all others previously examined on the current branch of search. At each iteration a new more attractive basis will be sought; if a basis does repeat itself, then no better solution than the current one can be obtained along that branch of search and a new 'attractive' branch must be found.

Initially it may appear that the necessity to store all previous computed bases would be limiting on the size of problem which can be solved but the string of variable values corresponding to each such solution can be stored as a binary digit. Obviously each basis is uniquely defined by an equivalent binary number and as one is interested solely in being able to distinguish between bases and not to investigate their actual form, a single number (converted to whatever base required) will represent the total storage requirement for one basis.

Termination will occur when either all goals have been met exactly or no further progress is admitted. Progress is prevented by the lack of any attractive variable to be introduced.

THE FORMAL ALGORITHM

Let
- $\mathbf{A}$ be the matrix of goal coefficients
- $\mathbf{w}$ be the vector of goal weights
- $\mathbf{b}^t$ be the vector of right hand side values at iteration t
- P be the set of previous bases on the current branch of search represented by a unique number p_t at iteration t
- P' be the set of all other bases previously encountered
- $\mathbf{r}$ be the vector of incoming variables producing new bases
- $\hat{X}_t$ be the set of variables in the basis at iteration t
- J be the set of variables to be considered
- I be the set of constraints to be considered
- J_k be the set of variables for which $a_{kj} < 0 \, \forall_j$

Vol. 6
1978

Calculate the best objective function value so far:

$$0^* = -\mathbf{w}\cdot\mathbf{b}^* \quad \text{where} \quad \mathbf{b}^* = \mathbf{b}^0$$

Set $\hat{X}_0 = \hat{X}^* = \phi$; $P = \phi$; $r = \{0,0,0\ldots0\}$; $t = 1$

Step 1. Set $J = \{1,2,3,\ldots,n\}$
$I = \{1,2,3,\ldots,m\}$
$j^{*1} = 0$

Step 2. If $I = \phi$ and $j^{*1} = 0$ then terminate with optimal solution of $\hat{X}^*$
$I = \phi$ and $j^{*1} \neq 0$ set $j^* = j^{*1}$; and go to step 6

Otherwise select goal i^* such that

$$d_{i^*} = \max_{i\in I}(d_i = \mathbf{w}_i\cdot\mathbf{b}'_i)$$

Break ties by selecting lowest value of i
If $d_{i^*} = 0$ and $j^{*1} = 0$ then terminate with optimal solution of $\hat{X}^*$
If $d_{i^*} = 0$ and $j^{*1} \neq 0$ set $j^* = j^{*1}$; and go to Step 6
Otherwise go to Step 3

Step 3. If $a_{i^*j} \leq 0$ remove j from J
Select variable j^* such that:

$$e_{j^*} = \min_{j\in J}(b'_{i^*} - a_{i^*j})$$

Break ties by selecting j^* such that:

$$z_{j^*} = \min.\ (z_j = \mathbf{a}_{j^1}\cdot\mathbf{w})$$

where $a^1_{ij} = \begin{cases} a_{ij}\ \mathbf{b}'_i \neq 0 \\ 0 \text{ otherwise} \end{cases}$

Break ties by selecting lowest value of j
If $z_{j^*} \geq 0$ go to Step 5
If $p_{t-1} \cup \{j^*\} \notin P$ go to Step 4

where $\bar{j}^* \begin{cases} j^* \ j^* \notin \hat{X}_{t-1} \\ -j^* \text{ otherwise} \end{cases}$

Otherwise set $j^{*1} = \mathbf{r}_{t-1}$
If $|\mathbf{b}'_{i^*} - a_{i^*j^{*1}}| < \mathbf{b}'_{i^*}$ go to Step 5

If $\sum_{j\in J_{i^*}} a_{i^*j} < b^*_{i^*} - |b'_{i^*} - a_{i^*j^{*1}}|$ and $\exists\, j \in J_{i^*}$

such that $a_{i^*j} < |\mathbf{b}'_{i^*} - a_{i^*j^{*1}}| + \mathbf{b}^*_{i^*}$ go to Step 5
Otherwise set $j^{*1} = 0$ and go to Step 5

Step 4. If $\mathbf{b}'_{i^*} - a_{i^*j^*} \geq \mathbf{b}_{i^*}$ and $J/j^* \subset P \cup P^1$ go to Step 5
If $|\mathbf{b}'_{i^*} - a_{i^*j^*}| < \mathbf{b}'_{i^*}$ go to Step 6
If $\sum_{j\in J_{i^*}} a_{i^*j} < b^*_{i^*} - |b'_{i^*} - a_{i^*j^*}|$ and $\exists\, j \in J_{i^*}$

such that $a_{i^*j} < |\mathbf{b}'_{i^*} - a_{i^*j^*}| + \mathbf{b}^*_{i^*}$ go to Step 6
Otherwise go to Step 5

Step 5. Remove j^* from J
If $J = \phi$ remove i^* from I, set $J = \{1,2,3,\ldots,n\}$ and go to Step 2
Otherwise go to Step 3

Step 6. $\hat{X}_t = \hat{X}_{t-1} \cup \{\bar{j}^*\}$
If $j^* = j^{*1}$ $P = P\backslash p_t$; $P^1 = P^1 \cup p_t$; $\mathbf{r}_t = 0$; $t = t - 1$
Otherwise $P = P \cup p_t$; $\mathbf{r}_t = j^*$
Let $\mathbf{b}^{t+1} = \mathbf{b}^t - \mathbf{a}_{j^*}$
Let $\mathbf{a}_{j^*} = -\mathbf{a}_{j^*}$
If any $\mathbf{b}^{t+1}_i < 0$, say i^1
Let $a_{i^1j} = -a_{i^1j}$
$b^{t+1}_{i^1} = -b^{t+1}_{i^1}$.
If $0 = \mathbf{w}\cdot\mathbf{b}^{t+1} < 0^*$ Let $0^* = 0$; $\hat{X}^* = \hat{X}_t$; $\mathbf{b}^* = \mathbf{b}^{t+1}$
Let $t = t + 1$ and go to Step 1

AN EXAMPLE PROBLEM

Minimize $\sum -w_i\cdot|d_i|$ where $w_i = -1$

s.t. $Ax + d = b$

where $\mathbf{A} = \begin{bmatrix} 1 & 3 & -5 & 1 & 6 \\ 2 & 6 & -3 & 2 & 2 \\ 0 & -2 & 2 & 2 & -1 \end{bmatrix}$ $\mathbf{b} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}$

Iteration 1.

Step 1. $J = \{1,2,3,4,5\}$; $I = \{1,2,3\}$; $0^* = 6$; $\hat{X}^* = \{0\}$; $p = \phi$; $\mathbf{r} = \{0,0,\ldots0\}$; $t = 1$; $\mathbf{j}^{*1} = 0$

Step 2. d_{i^*} = when $i = 2$

Step 3. min $(b'_{i^*} = a_{i^*j}) = 2$ when $j = 1,2,4,5$
min $z_j = -4$ when $j = 2$ and 4 $\therefore$ select $j = 2$ as lowest $j = 1,2,4,5$

Step 6. $\hat{X}_1 = \{2\}$
$p_1 = \{2\}$
$\mathbf{b}^2 = [0\ \ 4\ \ 2]^T - [3\ \ 6\ -2]^T$
$= [-3\ -2\ \ 4]^T$
$\mathbf{a}_2 = -[3\ \ 6\ -2]^T = [-3\ -6\ \ 2]^T$
$b^2_1 < 0$ and $b^2_2 < 0$

$\therefore$ $\mathbf{A} = \begin{bmatrix} -1 & 3 & 5 & -1 & -6 \\ -2 & 6 & 3 & -2 & -2 \\ 0 & 2 & 2 & 2 & -1 \end{bmatrix}$ $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$

$0 = 9 > 0^*$

Iteration 2.

Step 2. $d_{i^*} = 4$ when $i = 3$

Step 3. $J = \{1,3,4,5\}$
min $(b^2_{i^*} - a_{i^*j}) =$ when $j = 3$ or 4
min $z_j = -10$ when $j = 3$ $j = 3,4$

Step 6. $\hat{X}_2 = \{2,3\}$
$p_2 = \{2,3\}$
$b^3 = [3\ \ 2\ \ 4]^T - [5\ \ 3\ \ 2]^T$
$= [-2\ -1\ \ 2]^T$
$\mathbf{a}_3 = -[5\ \ 3\ \ 2]^T = [-5\ -3\ -2]^T$
$b^3_i < 0\ \forall_i$

$\therefore$ $\mathbf{A} = \begin{bmatrix} 1 & -3 & 5 & 1 & 6 \\ 2 & -6 & 3 & 2 & 2 \\ 0 & 2 & -2 & 2 & -1 \end{bmatrix}$ $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$

$0 = 5 < 0^*$ $0^* = 5$; $\hat{X}^* = \{2,3\}$

Iteration 3.

Step 2. $d_{i^*} = 2$ when $i = 1$ or 3 $\therefore$ select $i = 1$ as lowest

Step 3. $J = \{1,2,4,5\}$
min $(b^3_{i^*} - a_{i^*j}) = 1$ when $j = 1$ or 4
min $z_j = -5$ when $j = 4$ $j = 1,4$

Step 6. $\hat{X}_3 = \{2,3,4\}$
$p_3 = \{2,3,4\}$
$\mathbf{b}^4 = [2\ \ 1\ \ 2]^T - [1\ \ 2\ \ 2]^T$
$= [1\ -1\ \ 0]^T$
$\mathbf{a}_4 = -[1\ \ 2\ \ 2]^T = [-1\ -2\ -2]^t$
$b^4_2 < 0$

$\therefore$ $\mathbf{A} = \begin{bmatrix} 1 & -3 & 5 & -1 & 6 \\ -2 & 6 & -3 & 2 & -2 \\ 0 & 2 & -2 & -2 & -1 \end{bmatrix}$ $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$0 = 2 < 0^*$ $\therefore$ $0^* = 2$; $\hat{X}^* = \{2,3,4\}$

and so on, re-applying the algorithm.

Iteration 4. $\hat{X}_4 = \{2,4\}$

$$\mathbf{A} = \begin{bmatrix} -1 & 3 & 5 & 1 & -6 \\ -2 & 6 & 3 & 2 & -2 \\ 0 & 2 & 2 & -2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}$$

$0 = 10 > 0^*$

Iteration 5. $\hat{X}_5 = \{4\}$

$$\mathbf{A} = \begin{bmatrix} -1 & -3 & 5 & 1 & -6 \\ 2 & 6 & -3 & -2 & 2 \\ 0 & -2 & 2 & -2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

$0 = 3 > 0^*$

Iteration 6. $\hat{X}_6 = \{1,4\}$

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 5 & 1 & -6 \\ -2 & 6 & -3 & -2 & 2 \\ 0 & -2 & 2 & -2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

$0 = 2 = 0^*$ *Alternative 'optima'*

Iteration 7. $\hat{X}_7 = \{1\}$

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 5 & -1 & -6 \\ -2 & 6 & -3 & 2 & 2 \\ 0 & -2 & 2 & 2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$0 = 5 > 0^*$

Iteration 8. $\hat{X}_8 = \{1,2\}$

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5 & -1 & -6 \\ -2 & 6 & 3 & -2 & -2 \\ & 2 & 2 & 2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$$

$0 = 12 > 0^*$

Iteration 9. $\hat{X}_9 = \{1,2,3\}$

$$\mathbf{A} = \begin{bmatrix} -1 & -3 & 5 & 1 & 6 \\ 2 & 6 & -3 & -2 & -2 \\ 0 & 2 & -2 & 2 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$0 = 4 > 0^*$

Iteration 10. $\hat{X}_{10} = \{1,2,3,4\}$

$$\mathbf{A} = \begin{bmatrix} -1 & -3 & 5 & -1 & 6 \\ 2 & 6 & -3 & 2 & -2 \\ 0 & -2 & 2 & 2 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

$0 = 3 > 0^*$

Iteration 11. $\hat{X}_{11} = \{1,3,4\}$

$$\mathbf{A} = \begin{bmatrix} -1 & 3 & 5 & -1 & 6 \\ -2 & 6 & 3 & -2 & 2 \\ 0 & 2 & 2 & 2 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}$$

$0 = 8 > 0^*$

Iteration 12. $\hat{X}_{10} = \{1,3\}$

$$\mathbf{A} = \begin{bmatrix} -1 & 3 & 5 & 1 & 6 \\ -2 & 6 & 3 & 2 & 2 \\ 0 & -2 & -2 & 2 & -1 \end{bmatrix} \quad \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}$$

$0 = 9 > 0^*$

At this stage no further iterations can be made due to all possible 'improving' variables being excluded from consideration.

$\therefore$ alternative optima of

$$x_1 = x_4 = 1 \; x_2 = x_3 = x_5 = 0$$

and

$$x_2 = x_3 = x_4 = 1 \; x_1 = x_5 = 0$$

It may often be the case that rather than each variable being individually $0 - 1$ valued they are grouped in ordered sets as described by Healy [3] and Beale [1]. In Beale's terminology the problem is now characterised as being of *S*1 type, in that not more than one member of any set may take on a non-zero value in the final solution.

$$\text{i.e. } \sum_{i=1}^{k_1} x_i = 1, \; \sum_{i=k_1+1}^{k_2} x_i = 1, \ldots, \; \sum_{i=k_{n-1}+1}^{k_n} x_i = 1.$$

For this form of problem the algorithm can be readily adapted. Instead of the first feasible solution being a null basis it is necessary to produce a first feasible solution which complies with the *S*1 type constraints. This is simply achieved by selecting one variable from each ordered set according to the selection rules of the algorithm. The only other necessary change is at Step 6 where a_{ij^*} must be also subtracted from the coefficients of all variables in the same ordered set as the incoming variable.

Another extension which is, in fact, included in the present operating programme is the idea of range programming [see 5]. This is in recognition of the fact that the goals included in many problems cannot be stipulated exactly, but rather a range of acceptable outcomes can be given. The effect of this is simply to penalise attainment levels of the goals which lie outside this region of acceptability rather than to penalise values differing from a mean figure lying within this range. More generally the concept of a satisficing range can be adapted for use in the goal programming situation such that the deviation metric can be held more realistically to represent the reality of the decision maker's satisficing rather than optimising ideas. This concept can be further developed so that only over or under achievements are allowed. One of the basic ideas which the current algorithm exploits is the fact that deviations in either direction are equally undesirable but it would add little to the size or complexity of the algorithm to differentially penalise deviations of either one sign or the other. The inclusion of two separate weights for positive and negative types of deviations may also be included with no significant increase in complexity.

As indicated all variations presently available in existing $0 - 1$ goal adapted algorithms are possible with this new algorithm as well as the extension in the form of range programming. The size of problem which can be handled is also far superior. To date the largest problem solved is one connected with student nurse allocation which involved 148 goals, 296 variables and 24 *S*1 ordered set type constraints. The algorithm has been programmed in Fortran IV to run on a C.D.C. 7600 at the University of Manchester Regional Computer Centre and the problem size quoted produced an optimal solution in 106.18 CP seconds. Further computational experience is being gained at the moment.

Vol. 6
1978

REFERENCES

1. BEALE EML (1968) *Mathematical Programming in Practice*. Pitman, London.
2. GARROD NW, MOORES B and BRIGGS GH (1978) A zero-one goal programming formulation of the student nurse allocation problem. *Omega* **6**(1), 93–96.
3. HEALY WC (1964) Multiple choice programming. *Op. Res.* **12**, 122–138.
4. IGNIZIO JP (1976) *Goal Programming and Extensions*. Lexington Books, London.
5. LAURENT G (1976) A note on range programming: introducing a 'satisficing range' in a linear program. *Mgmt. Sci.* **22**(6), 713–716.
6. LEE SM (1976) *Linear Optimization for Management*. Petrocelli Charter, New York.

NW Garrod
B Moores
(*February* 1978)

Department of Agricultural Economics
University College of Wales
Penglais
Aberystwyth
Tysed
UK

l. 6
978

A Dynamic Programming Approach to Work Force Scheduling

FLOWERS and Preston [1] reported the results of a simulation study of the application of the search decision rule (SDR) to a work force scheduling problem for a manufacturing firm. The results indicated that the SDR methodology was easily implemented, did not require restrictive mathematical assumptions, was computationally efficient and produced superior results over the simulated time period (65 weeks) when compared to the actual company decisions. The purpose of this note is to demonstrate that the same work force scheduling problem may be formulated and solved optimally as a dynamic programming problem, and to compare its solution by dynamic programming to the SDR methodology. In this case, as is often true in management science applications, there are alternate ways to formulate and solve the same problem.

The special nature of this scheduling problem resulted in restricting the decision variable to include only the work force size in each period. These decision variables become the state variables for the dynamic programming formulation. To determine the possible values of the state variables for the specified thirteen week planning horizon, the maximum and minimum number of hours scheduled during the planning horizon are first determined. Then the integer number of employees necessary to allow the maximum scheduled hours to be worked completely on regular time is computed by dividing the maximum scheduled hours by the 37.75 net productive hours per week per employee and the result rounded up to the next integer value. The minimum work force size is determined analogously using the minimum scheduled hours and rounding the result to the next lower integer. The minimum and maximun work force size so determined are then compared to the beginning work force size for the planning period. If the beginning value is larger than the computed maximum, the maximum is reset to the beginning value. If the minimum computed value is greater than the beginning work force size, the minimum is reset to equal the beginning work force size. The resulting limits on the decision variable are used for each period in the planning horizon to determine the number of states considered. The same objective function as was used in the SDR analysis is also used for the dynamic programming solution. Standard backward induction procedures for dynamic programming problems were programmed for the same computer (IBM 370/145) that was used to solve the original problems with the SDR methodology, and the analogous fourteen runs for the thirteen week planning horizon were made. The results achieved by the dynamic programming algorithm included a total cost for the 65 week period of $194,004.71 as compared to the SDR value of $193,799.23 and an average execution time of 1.5 seconds per run for dynamic programming as compared to 11 seconds per run of the SDR procedure. The explanation of these results is relatively straightforward. The SDR methodology, while not optimizing each 13 week period, nonetheless finds 'frozen' solutions for four periods at a time (see Flowers and Preston [1]) which result in a total cost 0.1 per cent lower than the 13 week dynamic programming algorithm. The true optimum solution for the problem when solved as a 65 week dynamic programming problem results in a total cost of $193,210.81.

The computational results for the dynamic programming result in a clear advantage as compared to the SDR method. The SDR tries a number of values in its searches that are not promising (i.e. values above the maximum and below the minimum work force size) and the SDR has a step size used to change the search variable values which must be initially set to a fairly large value to explore a substantial area of the response surface and then later reduced when no further improvement with that step size is possible. This results in a larger number of function evaluations than are necessary for the implicit enumeration approach of dynamic programming.

REFERENCES

1. FLOWERS AD and PRESTON SE (1977) Work force scheduling with the search decision rule. *Omega* 5(4), 473–479.

A Dale Flowers
David L Dukes
Walter D Curtiss

Department of Operations Research
Case Western Reserve University
Cleveland
Ohio 44106
USA

(*April* 1978)

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 4, pp. 379–381, 1978. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

Brown G (1974) *French: L'informatique de décision en France.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Chanoine P (1974) *French: Le comportement des cours boursiers: 'trends' ou 'random walks'? Une application des chaines de Markov.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Chowdhury IG (1975–76) Job Sequencing with Single and Multiple Operations. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Dufer J (1974) *French: Le journal Le Devoir: cas de politique générale d'entreprise.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Dussart C (1975) *French: Les ensembles de marques de référence: une étude empirique sur leur existence et leur magnitude dans le processus de choix d'une marque.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Eilon A (1975–76) Aspects of Linear Programming and Accounting in a Manufacturing Company. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Faucher J-C (1975) *French: Lancement d'une marque privée de vodka.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Feldmann J (1973–74) *German: Wirtschaftssystem und Wirtschaftswachstum—Zur Systematik der Wirtschaftssysteme und ihrer Favorisierung der Determinanten des Wirtschaftswachstums.* PhD Thesis, George-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Flick R (1973–74) *German: Die Fluktuation selbständiger Existenzen im Handwerk.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Henze A (1972–73) *German: Die Sicherstellung der nationalen Güterversorgung. Allgemeine und spezielle Analyse am Beispiel der Versorgung mit Nahrungsmitteln in der BRD.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Hiersemenzel U-L (1973–74) *German: Die Rolle der Exporte in der wirtschaftlichen Entwicklung Brasiliens 1889–1964.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Kuhn H (1972–73) *German: Politischer, wirtschaftlicher und sozialer Wandel in Worms 1798–1866 unter besonderer Berücksichtigung der Veränderungen in der Bestellung, den Funktionen und der Zusammensetzung der Gemeindevertretung.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Mahlmann A (1973–74) *German: Standortrelevanz agrarpolitischer Massnahmen.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Mailloux J-B (1974) *French: Méthodes d'optimisation des estimateurs multivariés.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Meade N (1974–75) Applications of Management Science in Assessment of Venture Capital. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Mensah A (1972–73) *German: Untersuchungen zur Bankenstruktur ausgewählter Länder Afrikas.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Morin J-M (1974) *French: Influence des tests de produit sur le comportement d'achat du consommateur.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Phocas-Cosmetatos G (1974–75) A Principle for Approximate Solutions of Queueing Systems. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Rohlfing G (1972–73) *German: Der Entscheidungsprozess in der öffentlichen Verwaltung aus betriebswirtschaftlicher Sicht.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersäschen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

SCHINKE R (1972–73) *German: Der Zahlungsbilanzausgleich im Gemeinsamen Markt Zentralamerikas—Ein Diskussionsbeitrag zur monetären Integration von Entwicklungsländern.* PhD Thesis. Georg-August-Universität, Göttingen. (Available at: Niedersäschen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

VIELLARD E (1974) *French: La croissance des entreprises en franchising.* MSc Thesis. Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

VON OETTINGEN D (1972–73) *German: Externe Effekte und Wirtschaftsordnungspolitik.* PhD Thesis. Georg-August-Universität, Göttingen. (Available at: Niedersäschen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

ZINS M (1974) *French: La psychographie comme base de segmentation: Etude des relations entre profils généraux et spécifiques, une application.* MSc Thesis. Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

The following are obtainable at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, UK:

ADAMS J (1975) Short Term Forecasting. MSc Thesis. University of Bradford.

BIRCHALL DW (1975) Manual Jobs and Worker Attitudes and Behaviour. PhD Thesis. University of Bradford.

BIRRELL WD (1974) The Sequencing and Scheduling of Work Units with Manpower Resource Constraints. MSc Thesis. University of Bradford.

BURNETT JW (1974) Nature and Causes of Ego-Involvement in the Task. MSc Thesis. University of Bradford.

CHACKO WJ (1974) The Flow Shop Scheduling Problem. MSc Thesis. University of Bradford.

COCKS GB (1974) The Analysis of Portfolio Performance. MSc Thesis. University of Bradford.

CROWTHER G (1974) The Social Marketing of New Towns. MSc Thesis. University of Bradford.

ELLIS MED (1975) Management Potential as a Concept in Personnel Management. MSc Thesis. University of Bradford.

ELLIS GF (1975) Hypermarkets in France: Toulouse an Example. MSc Thesis. University of Bradford.

ELSAYED AM (1974) Marketing with Specific Reference to Pricing and Promotion in the Carpet Manufacturing Industry. PhD Thesis. University of Bradford.

FARRINGTON B (1975) Industrial Source Management. MSc Thesis. University of Bradford.

FIELD D (1975) Control in Organisations. MSc Thesis. University of Bradford.

GOODWIN CTE (1974) An Analysis of Dividend Policy and its Implications for Corporate Management. MSc Thesis. University of Bradford.

GORDON LM (1974) The Management of Liquidity: Planning for Adequate Liquidity for Working Capital throughout an Operating Period. MSc Thesis. University of Bradford.

GRIFFITHS GJ (1975) Marketing in Banking. MSc Thesis. University of Bradford.

HEWAGAMA LR (1974) Bank Advances: A Study of British Banking Policy and Strategy, with Regard to its Advances. MSc Thesis. University of Bradford.

HOLT CA (1975) An Examination of the Study of European Finance in the UK. MSc Thesis. University of Bradford.

JACKSON DJ (1975) Cell System of Production. MSc Thesis. University of Bradford.

JOHNSON IP (1974) Export Marketing in a Medium Sized Engineering Company. MSc Thesis. University of Bradford.

JUDKINS PE (1975) Reward of Scientists in the Atomic Energy Project. MSc Thesis. University of Bradford.

KENNEDY SH (1974) Formation of Company Images. PhD Thesis. University of Bradford.

LEONG KF (1975) Reliability of a Complex System with Repair. MSc Thesis. University of Bradford.

LEVENE FR (1974) A Study to Evaluate the Importance of Price Perception in Determining Purchasing Behaviour on Branded and Unbranded Products in Unitary Shopping Situation. MSc Thesis. University of Bradford.

MCLEAY SJ (1974) The European Stock Exchanges and Economic Efficiency. MSc Thesis. University of Bradford.

MARSDEN PWF (1975) European Financial Instruments. MSc Thesis. University of Bradford.

MARSH K (1975) Computer Packages. MSc Thesis. University of Bradford.

MEIDAN A (1974) Nonmetric Multidimensional Analysis of Export. PhD Thesis. University of Bradford.

MINTY AM (1975) Social Responsibility as an Investment Criterion for Multi-National Enterprises Investing in Less Developed Countries. MSc Thesis. University of Bradford.

MULLIGAN RM (1974) The UK Electronics Industry. MSc Thesis. University of Bradford.

NEWMAN D (1973) Business Strategy in the Smaller Firm. MSc Thesis. University of Bradford.

The following are obtainable at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1 Canada.

Brooksbank RD (1976) Agricultural Land Resources and Land Use in an Urbanizing Region. MSc Thesis, Ontario Agricultural College.

Campbell JC (1975) Identification and Evaluation of Factors Facilitating or Impeding Farm Adjustments and Milk Supply for Typical Dairy Farm Situations in One Region of Southern Ontario. MSc Thesis, Ontario Agricultural College.

Daniel GA (1975) A Mathematical Programming Model for Optima Small Farm Organization in St. Vincent. MSc Thesis, Ontario Agricultural College.

Dapaah SK (1974) The Effects of Soils and Time on the Accuracy of Crop Identification in S.W. Ontario from 1974 Satellite Imagery. MSc Thesis, Ontario Agricultural College.

Gurpersaud (1974) Factors Affecting the Establishment of Viable Economic Farms in the Wauna Area of Guyana. MSc Thesis, Ontario Agricultural College.

Hedley HA (1974) An Analysis of the Spatial Organization of the Ontario Feed Manufacturing Industry. MSc Thesis, Ontario Agricultural College.

Kado K (1975) Factors Influencing Japan's Agricultural Import Liberalization. MSc Thesis, Ontario Agricultural College.

Kizito EB (1976) A Multi-Period Linear Programming Approach to a Macro Livestock Sector Analysis with Special Reference to Uganda. MSc Thesis, Ontario Agricultural College.

Kramar RE (1975) Ontario Acreage Response: An Econometric Analysis. MSc Thesis, Ontario Agricultural College.

Maicher GB (1976) An Analysis of Economic Opportunities for Farm Adjustments on Selected Dairy Farms Shipping Farm-Separated Cream and Industrial Milk in Cans. MSc Thesis, Ontario Agricultural College.

Okyere WA (1975) A Study of Personnel Management Practices and Problems on Ontario Dairy Farms. MSc Thesis, Ontario Agricultural College.

Pieri RG (1975) The North American–Japanese Pork Industries—An Economic Analysis. MSc Thesis, Ontario Agricultural College.

Rabiu MO (1974) Identification and Evaluation of Factors Facilitating or Impeding the Establishment of Viable Economic Farm Types on Arable Farm Settlements in Western Nigeria. MSc Thesis, Ontario Agricultural College.

Robertson GC (1975) An Ontario Farm Management Simulation Game. MSc Thesis, Ontario Agricultural College.

Smith WC (1974) A Differentiated Pricing Policy to Improve the Dairy Industry in Guyana. MSc Thesis, Ontario Agricultural College.

Sonola SO (1975) An Economic Analysis of the Integrated Development and Management of Bottomlands (Basfonds) and Associated Uplands in the High Forest Area of Western Nigeria. MSc Thesis, Ontario Agricultural College.

Stackhouse SJB (1975) Organizational Responses to Uncertainty Under Conditions of Subjective Risk Aversion on Beef Feedlot Farms in Ontario. MSc Thesis, Ontario Agricultural College.

Yeboah AO (1975) Stabilizing Effects and Economic Benefits of Alternative Pork Stabilization Programs. MSc Thesis, Ontario Agricultural College.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 5, 1978. Pergamon Press. Printed in Great Britain

EDITORIAL

Doctor, Heal Thyself

ONE OF the intriguing, and necessary, questions directed at management science is how to measure its performance. This is not just a matter of theoretical curiosity, in the sense that students of action research are interested in observing the impact of actions taken within given systems; it is an exercise in justification, for those who contemplate becoming students of management science, for those who teach it, for those who practice it and for those who spend money on it. Since MS (Management Science) aims at making a contribution to the decision making process, and since its application must specifically be designed to improve the performance of an enterprise or an administration system, it is natural that people should pose questions about the effectiveness of management science itself and particularly ask whether the investment in its activities is worthwhile.

Some research workers refer to this area as OR on OR (Operational Research on Operational Research), although this term has a wider connotation than a mere evaluation of effectiveness. The latter concentrates on past performance, and while it is an important ingredient in OR on OR, it does not address itself to the question of how to improve OR in the future, how to select the most appropriate methodology and how to implement effectively the conclusions of OR/MS projects. What we are concerned with here is not OR on OR in this wider context, but the more limited issue of measurement and performance evaluation.

Now, in simple economic terms the benefit of an OR project could be determined as the difference between the accrued benefits and costs, a 'before and after' comparison, in the same way that any other investment decision is assessed. Such benefits and costs, measured in financial terms, may relate to absolute cash inflows and outlays over a given time interval, or to discounted flows (particularly when the time horizon stretches over several years), or to the payback concept (namely the period of time over which the investment is fully repaid). Ostensibly, therefore, it should be possible to construct a profit and loss account for any given project, or indeed for a whole OR/MS group, and draw conclusions regarding their economic efficacy. And, indeed, we often do precisely that. Such exercises play an important role in trying to understand the rise and fall of OR/MS in particular companies or even in some industries, and these exercises also serve

to explain the portfolio of projects that certain OR/MS groups are obliged to maintain in their continuous battle for survival within a questioning environment.

This profit-and-loss-account mentality is becoming so widely ingrained that many management scientists would not dream of either proposing or undertaking a project without the expectation of a positive outcome: Rightly, it is an attitude of mind that we should applaud. It ensures that management scientists have their feet firmly planted on the ground, that they constantly be alert to the needs of the organization in which they work, and that they make a worthwhile contribution to its well-being. Management science must be concerned with practical consequences, and if a profit and loss account of its activities helps to underline this practical orientation, then it undoubtedly serves a useful purpose.

But there is a danger in overdoing it, of becoming so obsessed with a single and narrowly defined measure of performance, that it may cloud people's perspective, and even lead to erroneous conclusions. This is a danger that needs to be properly understood, not least by evaluators of management science, when they come to pass judgement on its effectiveness.

Vol. 6
1978

I am reminded, in this context, of a conversation I recently had with Dr Gene Woolsey, the editor-in-chief of *Interfaces*. The editorial policy of *Interfaces* is to promote the practical applications of OR/MS in business, industry, or Government operations, urging that "emphasis should be directed toward operational problems encountered in implementation, i.e. 'selling' the model or procedure to the 'customer' ". So far, so good. It specifies, further, that a lead article in the journal "will have an *absolute* requirement of either *provable* savings of cost or *provable* increase in profits to the enterprise" (the italics are as stated by the editor). And, indeed, when you read lead articles in *Interfaces*, you find that they are accompanied by letters from senior executives or auditors testifying to the level of savings or profits achieved. These independent testimonials serve, the editor claims, as inviolable evidence of solid financial achievements; they are there for the readers to see, and for all potential users of MS to be duly impressed.

But how valid are they? Dr Woolsey takes the view that if conventional cost accounting methods, which constitute a widely accepted language for conveying financial information, state that in a particular case a certain benefit has materialized, then it must have taken place. One is bound to sympathize with this point of view. I have no doubt that such language is most invaluable for promotional purposes, and that it can indeed be persuasive in internal political manoeuvring to enhance the status and influence of an OR/MS group. When you are in Rome, speak as the Romans do; that should ensure acceptance, or—at least—that you are listened to. But as management scientists

we have to maintain a critical mind; we should not always believe that the kind of profit and loss statements that were referred to earlier, even when they are well intentioned and most scrupulously prepared, are accurate, or even necessarily true.

I say this because measurement, particularly in the managerial field, rests on certain conventions and assumptions (some explicit and others implicit), which are in turn based on a narrow view of the environment and the time horizon. That such a view is bound to be restricted is perhaps obvious for pragmatic reasons, in that it cannot encompass the many ramifications that a given action in a system is likely to have outside the confines of the framework for which the measurement is intended. Furthermore, even within acceptable conventions (and this is equally true in the accounting field), there are many possible alternative interpretations of given data, and it is naive to suggest that only one unique picture can be painted to describe a particular environment.

Let me cite some familiar examples to illustrate what I have in mind. A company undertakes a study of the diversification of its products and concludes that some low-volume lines contribute little to profit, but are a serious drag on costs. It is often not difficult to show that elimination of the laggards can greatly improve the control of a more limited product range, and so reduce costs to more than compensate for the effect of losses in sales revenue. The net financial outcome may well look positive, in the short term, but the reaction of competitors (either because it takes time to develop, or because it is difficult to quantify) is often ignored. In the context of a financial evaluation of the production system, it may well be true that a product variety reduction exercise in a particular situation is successful, and a monetary figure can be put to substantiate such a statement; but if as a result of this exercise competitors are able to become more firmly entrenched in the market, then the emerging picture is decidedly less rosy.

The example of an inventory control policy, which attempts to strike a balance between holding and runout costs, is often encountered by management scientists. The crucial question that usually arises is about the cost parameters that should be taken into account in the modelling exercise. In one particular case that came to my attention some time ago, opinions were divided not only about the appropriate rate of interest that should be adopted for calculating the holding costs, but also about the numerical value that should be put on the runout costs. The manufacturing people were content to equate runout costs with the profit that the company had to forego by not being able to meet demand; the marketing department felt that the loss of goodwill, even though intangible in the short term, was very serious. The estimates for the runout costs given by the protagonists varied by a factor of ten to one. As for the cost per order, one view was

that an average cost was appropriate (computed as the total cost, of which manpower was the major component, divided by the number of orders generated), whereas another view maintained that within a very wide range of ordering frequency the number of people employed in the office could not be adjusted and had therefore to be regarded as a fixed element, leading to a substantially lower result for the cost per order. Needless to say, the outcome very much depended on the assumptions in question, and what one section thought was an improvement in performance, emanating from a given set of proposals, was regarded by another section as no improvement at all, perhaps even a deterioration. Now, the natural reaction to such cases is to argue that the protagonists must resolve their differences before the study can be completed, so that an agreed set of data can be used for the purpose, and that, while the management scientist cannot dictate to the warring factions, he can help to bring them together and effect a reconciliation. But the example does illustrate that a testimony of savings in such a case is not based on some revelation of unadulterated truth, but that it merely represents one point of view. Even when a consensus is painlessly arrived at as a basis for completing the cost–benefit analysis, it need not coincide with the view formed by an outside observer, with his own interpretation of the same data.

Vol. 6
1978

Or take the design of remuneration packages, or career promotion schemes, based on performance. One can cite examples galore, like the operator on the shop floor, who increases his output as intended by a wage incentive scheme, but abuses tools and machines, because their accelerated rate of wear is not (or cannot) be debited to him; the salesman, who is paid a commission on the volume of sales and hence concentrates on what is easy for him to sell, and thereby distorts sales targets and planned product-mixes; the bank manager, whose performance is judged by the value of loans to customers: he therefore lends indiscriminately and his mistakes come to light long after he has been promoted and is no longer employed in the same branch; the departmental production manager, who—because of prevailing transfer prices—proceeds to buy from outside sources cheaper goods and services and ignores the consequences on feeder or servicing departments.

The argument that may be advanced against all these examples is that the incentive schemes are faulty and that with greater sophistication they could be made to take account of the negative consequences alluded to: if the operator is penalized for causing an increase in maintenance cost or a shortening of the life of machines, if the salesman is given differential incentives to push certain products, if the bank manager is demoted or sacked for mistakes in lending decisions, if the production manager is debited with a surcharge for buying outside—then all these people will behave more responsibly. We have

to be aware, though, of the fact that not only is the required level of sophistication difficult to design and to administer, but that often certain consequences cannot be properly assessed at the design stage, and when they begin to materialize (some take a long time for their effect to be fully appreciated), it may be too late or quite impractical to do very much about it.

Thus, apart from the fundamental questions of what is meant by cost, and how one should reconcile the different interpretations that may be suggested, there is also the problem that management performance is almost invariably allied to the concept of sub-optimization, in that it centres on a part of the system and on a given time horizon, to the exclusion of other effects, whether they are quantifiable or not. It is often not difficult, therefore, for a manager or management scientist to descend on a sub-system and improve its performance in simple accounting terms, only to discover subsequently that other sub-systems have suffered as a result, or that short-term improvements have been achieved at the expense of the long term. A better measure of benefit would attempt to look at the system as a whole, and over a considerable time period, in order to verify that the net outcome remains positive and that it is not transient. We all know the difficulties and the many pitfalls that one is likely to encounter on such a course, but it is incumbent on us to try.

The fact that the performance of one part of the enterprise can be improved to the detriment of other parts raises some serious questions about consultant–client relationships and about the ethical dilemma that a management scientist sometimes has to face. If he is commissioned to conduct a study for a departmental manager in an enterprise, does he have a responsibility only to that manager, even though other people are likely to suffer as a result? And how does he square this responsibility with his social conscience?

I have discussed some of these issues elsewhere[1] and do not propose to repeat the discourse here, except perhaps to mention time-and-motion-study as an interesting example of a technique that has plainly led to a very substantial short-term gain in countless enterprises, only to encourage in its wake a growing union militancy and a certain degree of social discord. I am not suggesting, of course, that if a balance sheet of benefits and disbenefits of time-and-motion-study were to be drawn, we would inevitably conclude that it has been wholly negative; but in retrospect it must be said that in calculating the localized benefits, management has chosen, perhaps unwittingly, to ignore the after-effects, and that industry is now paying the price.

Now, those who are totally dedicated to the profit motive will argue that it is not the business of management scientists, or managers, to concern themselves with social consequences, that their objective

[1] See "Responsibility of the Management Scientist", (1975) Editorial, *Omega* 3(5), 513–517.

in life is simply to make money (the more and the quicker the better), that in a competitive world it is everyone for himself, and that the measure of success is best described by a profit and loss account. If there is another drop to be squeezed out of a given system, they will not hesitate to extract it.

Dr Woolsey probably regards himself as a member of this breed of thrusters. He has further set himself the mission of making OR/MS a widely acknowledged tool (even a part) of management, and of shaking the academics out of their state of complacent stupor in order to ensure that management science becomes what it is supposed to be, namely concerned with the real world, with decision making that affects real people, and with performance of real systems. This latter objective, I wholeheartedly support. I even understand why a missionary needs to use shock tactics and somewhat superficial yardsticks to drive his point home. But as professionals we should maintain a modicum of humility and perspective. We also owe it to our students, to industry, and certainly to ourselves, to search for more comprehensive methods of measuring performance, not only of industrial operations, but also of our own activities.

Vol. 6
1978

SAMUEL EILON
Chief Editor

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 5. pp. 389 398
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483 78 1001-0389$02.00 0

Compliance Costs — The Cost of Paying Tax

MICHAEL GODWIN

University of Bath, UK

(*Received March 1978; in revised form May 1978*)

Compliance costs are the costs which taxpayers and others incur in meeting obligations imposed under tax legislation. The following paper reviews the existing body of research into these costs, considers the methodological problems revealed by research, and assesses the significance of compliance costs to public policy.

DEFINITIONS AND DISTINCTIONS[1]

COSTS OF compliance with tax legislation may be defined as the costs which are incurred by taxpayers or by third parties, over and above the liability for tax, in meeting requirements of the tax system. Compliance costs are thus distinct from administrative costs—the costs incurred by the tax authorities—but each complements the other, and they are interchangeable to some extent. For example, under the American system of self-assessment to income tax, each taxpayer fills in a detailed annual return and calculates his own liability, whereas in the UK, the bulk of this work is done by the Department of Inland Revenue. Thus the burden falls to a greater extent on the income taxpayer in the United States, while more of this work is done by the administrators in the UK. Proponents of self-assessment argue that it makes manifest the way in which the citizen's tax bill is derived, and it reduces administrative costs. Opponents of this system point out that it can impose duties in an arbitrary and discriminatory way on differently-placed taxpayers —a centrally administered system is more consistent and less covert. The use of productive resources may be the same regardless of whether the work is done in the private or in the public sector.

The total of compliance plus administrative costs is defined as the operating cost of a tax. It seems eminently desirable that operating costs should be minimised. Taxpayers should not be compelled to perform unduly arduous tasks merely for the pleasure of paying tax, while revenue authorities should spend as little on administration as possible consistent with successful collection of tax.

Compliance with tax laws can involve collecting revenue, accounting for it, preparing tax returns, and remitting tax. This in turn requires an understanding of the relevant tax legislation. Costs may be incurred by private individuals, by businesses and other organisations such as universities and charitable bodies. They are not confined to direct cash expenditure. Unpaid time spent on tax compliance could be used in other ways, and may be valued in terms of the opportunity cost of activities foregone. There may also be a psychological burden arising from legislation which requires the individual to perform complex duties efficiently under sanction of a legal penalty. As Adam Smith put it "Though vexation is not strictly speaking expense, it is certainly equivalent to the expense at which every man would be willing to redeem himself from it" [20]. Compliance costs may be imposed as a structural feature of a tax—for example, where employers are made responsible for withholding employees' income taxes. These costs are a permanent feature of the tax, and they continue throughout its life. Other costs are associated with changes in the tax system and these costs tend to be very high at the introduction

[1] This section draws on Sandford [18].

of the tax and subsequently tend to decline gradually until they stabilise at the long term 'permanent' level. The once-only costs incurred at the introduction of a tax may be referred to as commencement costs, while costs above the permanent level which arise during the 'settling-down' process may be defined as temporary costs. A further complication is that no tax can be looked at in isolation—obligations imposed under a new tax must be balanced against the ending of costs which existed under superseded taxes.

A distinction may be drawn between the minimum costs legally imposed—the mandatory costs of the tax—and the additional costs which the individual may choose to incur in the pursuit of some other advantage, usually the reduction of tax liability. These are referred to as discretionary costs of the tax. Some commentators including Johnston [11] argue that only mandatory costs should be classified as pure costs of compliance with tax legislation. However, discretionary costs would not arise if tax laws did not exist, so it may be argued that they too are a foreseeable consequence of the legislation. In the UK, Selective Employment Tax, which was effectively a poll tax on employed labour in the service and construction industries, provides an illustrative case of a tax under which there was considerable scope for reducing tax liability by incurring discretionary compliance costs. Many employers were able to cut their SET bills by such expedients as rearranging the siting of service sector employees as far as possible into establishments where the majority of employees were exempt manufacturing workers. Under the legal definition of an establishment, *all* employees on such a site became exempt. Employees could also be encouraged to seek self-employed status and thereby avoid the tax.

RESEARCH METHODOLOGY AND FINDINGS

(a) *Corporate taxation*

The bulk of the studies of the compliance costs of corporate taxes have been undertaken in the USA, where pioneering research on the topic appeared in the mid-thirties. The earliest published study known to the author is by RW Haig [10], who mailed a questionnaire to 1600 members of the American Management Association in 1934, asking questions on the costs of complying with various state and federal taxes. He obtained a 10% usable response, mainly from larger corporations averaging $17 m sales per year.

Compliance costs ranged from 1.2% of tax liability for property tax up to 9.5% for state and local income tax, with an average of 2.3% of total tax liability for all taxes. Haig found that, with income taxes, high compliance costs were associated with low administrative costs —while sales and property taxes had lower compliance costs and higher administrative costs. He therefore asserted that there was a trade-off between compliance and administrative costs.

The number of states in which a firm traded also influenced the level of compliance costs. This finding has been confirmed in subsequent American research [13, 15, 17]. The general validity of the findings may nevertheless be queried, firstly because of the poor sample response rate, and secondly because sales and income tax liabilities were abnormally low in 1934, during the depths of the depression, distorting the ratios of compliance costs to tax liabilities.

A similar questionnaire was distributed in Canada to 500 corporate supporters of the Canadian Tax Foundation in 1961 by MH Bryden [3] who obtained a 25% response.

The findings of this survey differentiate between the "costs of paying taxes" for which the company itself is liable, and the "costs of collecting taxes" with regard to the tax liability of others.

"Costs of paying tax" as defined averaged 0.75 of one per cent of tax remitted, while "costs of collecting tax" averaged 0.5 of one per cent of taxes collected, of which half was attributable to deduction of personal income tax at source.

It was found that larger companies were able to mechanise many compliance procedures, and there appeared to be economies of scale in collecting taxes. The compliance costs of minor taxes were very high in relation to the amounts collected.

An interesting observation, which seems to have been corroborated in many compliance cost studies, was that there was a wide range of reported costs even among taxpayers in

Vol. 6 1978

apparently similar situations. Ratios of compliance costs to tax remitted ranged from 0.04% to 9.76%! Although economies of scale partly explain this, no fully satisfactory explanation of this phenomenon has yet been offered; the present author suggests that the following factors may be important:

(i) *Ability, education and training.* Individual taxpayers obviously vary in their ability to deal with tax and some businessmen may have the bookkeeping training and ability to cope easily with tax affairs, while other successful entrepreneurs may not have such skills, and lack even the background knowledge to delegate such work effectively.

(ii) *Attitude to taxation.* It seems possible that some taxpayers are unusually hostile to governmental intervention and tax laws, and such taxpayers will go to unusual lengths to minimise tax liability, spending considerable amounts of time and money in the process. They thus incur high "discretionary" costs.

(iii) *Valuation of time.* An alternative hypothesis is that some taxpayers place an unusually low value on time, and are prepared to spend a considerable amount of it if they think there is some prospect of reducing the amount of money paid out in tax.

Bryden recognised that his sample frame was not drawn randomly, and again the response rate was low.

The weaknesses of the mail questionnaire approach are, firstly, that response tends to be poor, and secondly, that the reported findings have not been verified empirically by the researcher. One alternative approach is that of personal interviews, during which points of detail can be probed more fully in order to obtain more accurate findings. The disadvantage of this approach is that it can be extremely time-consuming, so the researcher is often limited to a small sample or to a particular tax, or both. An example of this type of approach is KS Johnston's study of six Columbus, Ohio, manufacturing corporations' costs of complying with federal income tax legislation [11]. He defines the compliance costs of this tax as "the reduction in the corporation's operating costs, exclusive of the tax itself, which would result if the federal income tax were eliminated". The most interesting finding of this study was that when compliance costs, expressed as a percentage of taxable income, were analysed by size of firm, the results appeared to provide clear evidence of economies of scale.

A good example of a detailed study of the compliance costs imposed under a new tax is Oster & Lynn's work on the Ohio Axle Mile Tax [17]. They selected 18 trucking companies who were liable to this tax according to the Ohio mileage of trucks (lorries), at rates which increase with the number of vehicle axles. These companies were chosen to represent different sizes and types of firm and locations within the state. Eleven of them co-operated in developing estimates of compliance costs, which included collecting information, preparing returns, payment of the correct tax, obtaining judicial and administrative interpretation of the law, and adjusting operations to minimise tax liability. Oster and Lynn thus include "discretionary costs" in their analysis.

The main costs identified were incurred by drivers; clerks and terminal despatch staff were also required to perform substantial duties. Overall, compliance costs were almost 19% of net tax.

Compliance costs were again found to be highly variable. An important factor was the predictability of operations—where vehicles of known tax rating worked regular routes of known length, compliance costs were minimised. It was also observed that com-

RESULTS OF JOHNSTON'S STUDY

Firm	Size	Discretionary costs ($)	Total compliance costs ($)	Compliance costs as a percentage of taxable income
A	Large	5,965	51,918	0.10
B	Large	36,573	193,111	0.20
C	Medium	—	23,039	0.41
D	Medium	8,150	37,312	0.56
E	Small	—	3,817	3.24
F	Small	524	1,511	2.95

pliance expenses are incurred even where tax liability is low—a company running trucks mainly outside Ohio reported a low taxable mileage within the state, but compliance duties involved in the repeated state border crossings were numerous.

American research repeatedly refers to the substantial increases in compliance costs resulting from inter-state trading. Although this has not been a particular problem in Europe to date, such difficulties may well become general within the EEC unless efforts towards harmonisation of European taxes are pursued vigorously.

It is also worth noting that if continued pressure towards devolution does eventually result in the inauguration of independent parliaments within the United Kingdom with powers to impose taxation and direct public expenditure, these problems could well arise within the UK in the foreseeable future.

(b) *Other business taxation*

There have been three studies of American sales taxation, the largest by Professor Yocum [25], who examined the sales tax collection costs of 526 Ohio stores having gross sales of $50,000 and over for July–December 1959. Costs incurred in the computing of tax liability by clerks at the time of sale were examined by time study methods—6,700 sales transactions were timed by stopwatch and the time spent was valued accordingly. Other data were collected by *interview*, a team of 25 graduate accountants carrying out all interviews and time studies. It is apparent that this must have been one of the most comprehensive and costliest compliance cost research projects undertaken.

Costs were divided into administrative, sales clerk, and overhead expenditures. Administrative costs averaged 5.55%, and sales clerk costs 2.38% of tax liability, and if a portion of overhead expenditure was also allocated to compliance, total collection costs were nearly 15% of liability. Yocum concluded that differences in sales tax collection costs "are in general associated with differences in ... size of store, in size of average taxable transaction, and in the ratio of taxable to gross sales ... costs as a percentage of tax liability are generally higher for smaller stores, lower for larger stores. This ... is probably associated with the fact of certain fixed elements in collection costs, and economies of scale in store operations".

These findings appear to be corroborated in other studies. All the evidence so far collected seems to indicate that the burden of compliance weighs particularly heavily on the small businessman and the self-employed individual.

Muller [16] conducted an extensive study during which 250 small businesses in Washington State were sent questionnaires; 198 responded, and interviews were conducted with 100 of these. Supplementary information was obtained from tax commissions and accounting firms.

The 100 interviewed respondents were asked to fill in "tax compliance time reports", and 75 provided usable data.

This questionnaire plus interview technique has the advantage of both using a large initial sample, and also permitting a deeper examination of particular cases. Muller emphasises that the emotional strains imposed on the small businessman by tax compliance duties are often considerable—he concludes that "the small businessman, pressed into the role of tax collector, is troubled about the uncertainty of his legal tax liability, the multiplicity of reporting and other requirements of the taxing authorities, investment in clerical and accounting services, and the time and emotional tension required for tax compliance. The problem is a very real one, and in some states the small businessman feels the choice is between compliance and failure, or non-compliance and survival. Time and again the investigators were told that the interviewee had no objection to paying taxes, but the methods of record-keeping and reporting had become so complicated that it grew into an unreasonable burden".

These results are very similar to some findings from West Germany in Strumpel's survey of German small businessmen [21]. He obtained data from a probability sample of 1009 self-employed businessmen and professionals, and concluded that the high psychological burden of the requirements of tax legislation had created a climate of hostility to the tax authorities, in which evasion was condoned, and the self-employed were prepared to incur considerable costs in time and money in order to reduce tax bills, feeling that the burden of compliance amounted to a bias against the small business. Compliance costs

Vol. 6
1978

were again found to be regressive, bearing most heavily on the lower-income-bracket respondents.

Two small studies of Value-Added Tax compliance costs have been undertaken to date. VAT differs from the typical American state sales tax in several important respects:

(i) Although the total tax collected is equal to the yield of a single-stage retail sales tax at the same rate, VAT is collected in instalments at all stages of production. Compliance costs are imposed on primary producers, manufacturers, wholesalers and all other traders; under a retail sales tax, only the retailer suffers substantial compliance burdens. Thus an examination of retailers' VAT compliance costs will reveal only part of the total compliance costs of the tax.

(ii) American sales taxes are normally computed and applied at the time of sale, so sales clerk time is an important element in compliance costs. In the UK, retailers generally calculate tax-inclusive prices on supplies before placing them on sale.

(iii) A single stage tax only involves the return of sales records, whereas under VAT, information about both sales and purchases must be given.

The first study, by Barker [1], investigated the effects of introducing a model VAT at a single rate, with no exemptions, on six American firms. It was found that under these assumptions, the number of sales and purchases invoices processed determined the level of compliance costs. The number of invoices is of course also linked to the average transaction size. However, these results were to a large extent predetermined by the simplifying assumptions of the model tax.

The present author has conducted a study of the compliance costs imposed on Bath retailers at the introduction of VAT in the UK [8]. Twenty-nine retailers were each interviewed three times between June 1973 and October 1974. The re-interview technique was used in the hope that it would help to distinguish between permanent and temporary costs.

Principal findings from this survey were that compliance costs were highly variable, and their size was influenced by the rating of goods sold, and the "special scheme" for calculating VAT which was chosen. The change from the superseded purchase tax had been very costly and time-consuming, and further changes in the VAT were resented. The introduction of a higher rate in addition to the standard rate probably increased compliance costs substantially. Many interviewees selling largely zero-rated goods reported low or negative tax payments, and it was apparent that this group objected to their compliance duties more strongly than traders with a greater compliance burden who were nevertheless making substantial tax payments. The mean cost of time spent on VAT within these firms was estimated at £140 each for the financial year 1973/4, which is equivalent to the median gross earnings of a male manual employee for 3 weeks 2 days in that year.

(c) *Compliance costs of personal taxpayers*

JH Wicks has carried out several studies of state and federal income taxation in Montana, USA [22–24]. He concluded from questionnaires sent to parents of Montana students that compliance costs varied significantly by occupation, with the self-employed bearing the highest costs. Compliance costs were not related to income. It was noticeable that a few taxpayers in each of Wick's studies reported spending unusually large amounts of time on compliance, and there seems to be no simple explanation of this. Mean compliance costs are therefore often much higher than the median.

In the UK, the only study on personal taxpayers' compliance costs is Sandford's two-stage survey, which was supplemented by a survey of professional tax advisers [19]. The first stage of this survey was a sub-interview on two NOP Market Research Ltd surveys of statistically-selected electors. 2,773 respondents, a 78% response, provided basic information on time and money spent on tax affairs. Those who paid for tax advice, or spent more than 8 hours a year on tax matters, (designated "high compliance cost taxpayers") were then re-surveyed more extensively. It was concluded that in 1970, at least $8\frac{1}{2}\%$ of personal taxpayers employed paid advisers, and a further $3\frac{1}{2}\%$ used unpaid advisers. Taxpayers in England

COMPLIANCE COSTS OF FEDERAL INCOME TAXPAYERS BY OCCUPATION

Occupation	Cost in $		Cost as per cent of tax liability	
	Mean	Median	Mean	Median
Self-employed, non-professional (N = 33)	333	206	48.9	15.5
Professional (N = 22)	273	90	23.4	13.5
Managerial (N = 20)	153	69	13.5	6.0
Sales (N = 5)	130	63	34.6	13.0
Craftsmen (N = 9)	75	45	8.2	3.0
Other white collar (N = 9)	38	47	8.2	8.0
Other wage labour (N = 19)	30	26	4.9	4.0

Source: JH WICKS. Taxpayer compliance costs from personal income taxation. *Iowa Business Digest*, August 1966, 16–21.

and Wales spent at least 72 million hours on tax affairs themselves in that year. Fifty per cent of "high compliance cost taxpayers" were self-employed. Compliance costs were "inequitable, with a marked tendency towards regressiveness...the incidence of compliance costs is not evenly distributed throughout the population, nor is it directly related to income". Compliance costs were greater than administrative costs, and the combined total operating costs of personal taxation were at least 3.8% of tax revenue, and were possibly as high as 5.8%.

CONCLUSIONS FROM EXISTING RESEARCH

A tentative summary of some of the main conclusions from existing research would be:

(i) That the compliance costs of changes in the tax system are high.

(ii) That compliance costs are highly variable, tending to bear most heavily on those whose taxable activities are unpredictable.

(iii) That there are economies of scale to compliance—they are a greater burden on the small than on the large business.

(iv) That self-employed personal taxpayers incur high compliance costs.

(v) That the compliance costs of minor taxes are often high in relation to revenue.

(vi) That inter-state trading in the USA increases costs of compliance.

(vii) That sales tax compliance costs vary with the number of sales transactions and the number of tax rates.

METHODOLOGY—THE PROBLEMS REVEALED

(a) *Response*

Although a large number of studies of tax compliance costs have been published, few have been rigorous in their methodology. The earliest research projects were marred either by low response rates, or by a failure to specify the successful rate of response. Even a more recent and thorough project such as Yocum's study utilises a rather questionable sampling procedure, under which non-response was compensated for by the substitution of new firms. Consequently the magnitude of non-response is unknown, and the possibility of bias arising from differential response by different types of respondent cannot be discounted. It is also apparent that taxation is a sensitive subject for research, and all too often insufficient care taken in the preparatory stage has resulted in a poor response to a survey. In general the response rates to all but the smallest surveys are in the 30–45% range. Muller, and later Sandford, have obtained better response from two-stage surveys, in which a simple initial questionnaire is distributed to a large sample. Respondents of particular interest to the investigator are then re-surveyed by interview and/or a more detailed questionnaire. Crucial data are thus obtained simply from a large sample, while more detailed and open-ended questions are investigated at the second stage.

(b) *The sample frame and technique of investigation*

The technique of investigation is the major influence on the size of sample selected. In general, small surveys have been conducted by personal interviews, which permit a thorough examination of each individual case. The disadvantage of a small sample is of course that there is no indication of the representativeness of the results. With a larger sample, the time and cost of the interview technique are usually prohibitive. Yocum's survey is the only large study where extensive interviews and time-and-motion studies were carried out. The costs of employing 25 staff of an accounting firm to undertake this data collection for 526 retail stores must have been considerable. All the other large scale surveys have been carried out by mail questionnaire, which is in many ways less satisfactory. Response rates drop off considerably, and the possibility of questions being misinterpreted is much increased. Bad questionnaire design may result in the omission or distortion of certain data, whereas such problems are soon apparent in an interview, and can be cleared up quickly. The researcher investigating by mail questionnaire never actually meets respondents, and so cannot form any subjective assessment of their reliability and frankness.

Again the best solution appears to be a compromise utilising a large questionnaire sample and a smaller sample of interviewees. Certain investigators—Muller, Wicks and Killworth, and Sandford, have also made use of supplementary interviews with employers of respondents, government officials, tax advisers and other informed sources. It is evident that if compliance costs are to be examined comprehensively, a large scale survey procedure, with additional evidence drawn from several corroborating sources, may well be necessary. Since this will often be infeasible, it is unsurprising that many researchers have preferred to opt for detailed investigation of up to a dozen respondents.

The work of Matthews [14] and of Yocum is interesting for its use of time and motion study techniques in the examination of sales tax compliance costs. With the typical American sales tax, the amount of tax is calculated and added on at the time of sale, goods being priced tax-exclusive. It was established by Matthews, using stop-watch timings, that the cost of this procedure in additional sales clerk time was in excess of the costs of record and book-keeping for the tax. However, Yocum found that accounting costs were substantially greater than sales clerk costs. Where, as in these cases, the time spent on tax compliance can be accurately identified and timed, it should be possible to measure costs of compliance quite precisely.

(c) *Apportionment and allocation*

Usually, however, it is necessary to estimate the cost of compliance from less specific data about the respondent. If a taxpayer hires an accountant to unravel all his financial affairs including tax, the accountant will normally present him with an unitemised bill. Clearly some proportion of this bill represents a tax compliance cost, but how much? A useful methodological approach is to consider how different the situation would be in the absence of all taxes. If the accountant was hired primarily to cope with tax problems, the bulk of his fee may be classed as a compliance cost. If, however, the tax work forms a minimal proportion of the total, it may be unrealistic to attribute any of the fee to compliance. In either case, though, the conclusion reached will be subject to a margin of error.

There has been considerable discussion, particularly by Yocum, of the problem of allocating overheads. This hinges on the question of whether tax compliance necessitates an increase in overhead charges or whether compliance work simply results in a more intensive use of existing facilities. If a business has to rent, heat and light additional offices solely for its tax department, the costs of such offices are clearly compliance costs. However, in many cases tax work is done within an accounting department which would continue to be needed in the absence of taxation. In this case it may be argued convincingly that the overhead expenditure is independent of tax compliance, and the compliance cost is zero. It may be that the percentage of overhead expenditure allocable to tax compliance increases with the size of firm and the complexity of tax duties performed. The sole proprietor who stays at work for an extra hour every month to update his tax accounts and returns will incur minimal

extra costs in heating and lighting. A larger firm may find it efficient to set up a separate accounts department, renting additional space which will be devoted to tax work for part of the time. The largest businesses may well have offices maintained exclusively for tax accounting purposes.

The problem of discretionary and mandatory compliance costs is also difficult to resolve. A business may often allocate an amount of resources to tax accounting far above those necessary to comply with the law. Tax planning on a large scale is a normal feature of modern corporations. However, it has been observed that this expense, although tax-related, is not strictly compliance expenditure. The directors of the business choose to incur certain costs for reasons of their own, usually in order to reduce liability to tax. An attempt should be made by the researcher to isolate the mandatory element of compliance expenditure, and where this is difficult, it should be clearly stated that the estimate of compliance costs includes a "voluntary" element which is not legally imposed.

Another difficulty may arise in differentiating between temporary, permanent, and isolated costs of compliance. An observer may well overestimate the permanent level of compliance costs if he investigates them during a period when a new tax has just been introduced. Temporary problems of adjustment will be eliminated as the tax becomes familiar, and many "commencement" expenditures may occur once only, such as the heavy outlay on small calculator sales which accompanied the introduction of VAT in the UK. Again considerable care is needed in the allocation of costs to the correct category. Where a single business has incurred large legal fees in a tax dispute, this will usually be categorised by the researcher as an isolated case of high costs. If, however, it emerges that due, perhaps, to poorly-drafted legislation, a large percentage of taxpayers are in dispute with the authorities then such expenditure may be considered a regular cost of compliance.

(d) *The problem of valuing time*

Much of the cost of compliance arises in the form of time spent by personal taxpayers, accountants, office staff and others. It is evident that if we wish to measure total compliance costs in money terms, the value assigned to such time will be a principal determinant of the total estimated cost. If a realistic attempt to measure compliance costs is being made, it is important that every care should be taken to value time accurately. Unfortunately, there is widespread disagreement as to what the value of time really is [7, 9]. Where the time of employees is concerned, it seems appropriate to value it at the going rate for the relevant grade of staff. However, if salaried staff are included among overhead costs, many of the arguments set out above on the subject of overheads apply. In particular some commentators contend that additional tax work may be done during normal working hours by existing staff working harder. If this is the case, then it can be argued that the cost to the firm is nil. Where there is clear-cut evidence of extra staff having been taken on to do tax work, or of existing staff doing this work in paid overtime, the case for valuation at the wage rate, or even at the overtime rate, is much stronger. A more difficult problem arises in the case of the personal taxpayer or self-employed taxpayer who does tax work outside normal working hours.

Generally, I would consider that time spent on tax compliance by the self-employed should be valued at the occupational wage rate—though this may not be easy to ascertain for many of the self-employed. Strictly, there should also be a "risk" element added into the value, but it would be difficult to estimate. Valuation of non-work time spent by personal income taxpayers on compliance is more difficult. Perhaps where the fee paid to a tax adviser for doing similar work is known, this should be used as a basis for the costing.

(e) *Benefits from compliance*

An allowance should be made when measuring the costs of compliance for certain mitigating factors. Where a new tax is introduced, the duties imposed are often offset by the ending of obligations which were required under abolished taxes. Also, if businesses have to keep detailed records because of tax legislation, the result will often be that they maintain more effective control and operate more profitably than would otherwise have been the case.

THE SIGNIFICANCE OF COMPLIANCE COSTS

It is apparent that much tax legislation has

been brought in without careful consideration of the cost of compliance. Although governments are constantly under pressure to minimise the administrative costs of their operations, the same constraints do not apply to the imposition of unpaid duties on groups of taxpayers. The result is often that work which could be carried out efficiently by trained civil servants is performed less accurately and reliably by unqualified taxpayers, at perhaps a greater overall cost to the economy. It is regrettable that this issue is frequently ignored even in policy discussions where it is of crucial significance, as in the debate over the relative merits of self-assessment to income tax, as operated in the USA, and of assessment to income tax by the authorities, as operated in many European countries. For instance a new proposal by Barr, James and Prest [2] for a UK self-assessment system has no detailed consideration of the issue of relative efficiency and overall operating cost to the economy. It may also be pointed out that the Inland Revenue estimates the cost of administrative manpower to operate a self-assessed local income tax in the UK to be four or five times as much as if the tax were integrated with the national income tax [12]. Dean asserts that the idea that governmental costs of a self-assessed tax will be greater than for other types of income tax "flies in the face of all logic" [5]. There is an evident lack of appreciation of the facts both by the proponents and the antagonists of such a change.

Only in the United States has any great interest been shown in the problems of compliance. There seems to be little awareness of the economic implications of compliance costs in Europe. A major new tax, VAT, has been prescribed for use throughout the EEC without any apparent preliminary research on the probable costs of compliance, and, what is even more surprising, without any subsequent large-scale monitoring of these costs in any of the member countries. This might be comprehensible if the tax impinged on a small tax collecting population who were required to perform few duties, but in every case the population of registered traders is enormous (over one and a quarter million traders in the UK) and the record-keeping and assessment duties involved are considerable.

In the UK, many compliance duties are also imposed on employers under the PAYE system, but apart from unsubstantiated guesses that the compliance cost is about equal to the administrative cost of this institution, no published estimate of this cost has been made, though Dean [6] has suggested that compliance costs per employee are inversely related to number of employees, and hence tend to be regressive.

Compliance expenses can be incurred by beneficiaries of the fisc as well as by contributors. It is often necessary for those claiming social security benefits to incur considerable costs in fares, telephone calls and postage to establish their claims, and it appears that the psychological costs imposed on many claimants, particularly the old, by the processes of interviewing and form-filling, often deter those most in need from claiming their entitlements. In the UK at least the social security system seems ripe for reform. Colin Clark lists forty-four separate state means-tested benefits [4] administered by many different UK, national and local authorities. To claim the various benefits and allowances, applicants often have to provide the same information many times over to different bodies. A thorough investigation of this system might well reveal possibilities for rationalisation and simplification.

It may be concluded that the significance of the costs of compliance has been largely unrecognised in the past, and these costs are still frequently overlooked by policy-makers in weighing their decisions. This has resulted in an unsatisfactory position in which many individuals, particularly small businessmen and the self-employed, are having to contend with peculiarly onerous legislative requirements. It is however difficult to make many dogmatic assertions about the effects of the compliance burden since so little reliable evidence on the problem has yet been brought forward.

REFERENCES

1. Barker PA (1972) Value added tax, the cost to the businessman. *J. Accountancy*, **134**, 75–79.
2. Barr NA, James SR & Prest AR (1977) *Self-Assessment for Income Tax*. Heinemann, London.
3. Bryden MH (1961) *The Cost of Tax Compliance*. Canadian Tax Foundation, Toronto.
4. Clark C (1977) *Poverty before Politics*. Institute of Economic Affairs, London.
5. Dean PN (1977) The true tax cost. *Accountancy Age* **8**(13), 16–17.

6. Dean PN (1975) Some aspects of tax operating costs with particular reference to personal taxation in the United Kingdom. Unpublished Ph.D. Thesis, University of Bath, UK.
7. Department of Environment (1970) Papers and proceedings of a conference on research into the value of time. *Time research Note 16*, London.
8. Godwin MR (1976) VAT—compliance costs to the independent retailer. *Accountancy* **87**(997), 48–60.
9. Grazia S De (1964) *Of Time, Work and Leisure*. Doubleday, New York.
10. Haig RW (1935) The cost to business concerns of compliance with tax laws. *Mgnt Rev.* **24**, 323–333.
11. Johnston KS (1961) Corporations' federal income tax compliance costs. Ohio University Bureau of Business Research. Monograph 10, Columbus.
12. Layfield F (1976) Report of the committee of inquiry into local government finance, Cmd 6453, HMSO, London.
13. Mathes SM & Thompson GC (1950) The Tax on taxes. *Conf. Board Business Rec.* (April), 130–137.
14. Matthews MP (1956) A measurement of the cost of collecting sales tax monies in selected retail stores. University of Utah Bureau of Economic and Business Research, Salt Lake City, USA.
15. May JB & Thompson GC (1959) The high cost of tax compliance. *Bus. Rec.* (August) 383–388.
16. Muller FJ (1963) The burden of compliance. Seattle Bureau of Business Research, University of Washington.
17. Oster CV & Lynn AD (1955) Compliance costs and the Ohio axle mile tax. *Nat. Tax J.* **8**(2), 209–214.
18. Sandford CT (1976) Tax compliance costs matter: Chancellor please note. *British Tax R.* **4**, 205–212.
19. Sandford CT (1973) Hidden costs of taxation. Institute for Fiscal Studies, London.
20. Smith A (1776) *Wealth of Nations*. Book 5, Chap. 2, part 2.
21. Strumpel B (1966) The disguised tax burden. *Nat. Tax J.* **19**(1), 70–77.
22. Wicks JH & Killworth MN (1967) Administrative and compliance costs of state and local taxes. *Nat. Tax J.* **20**(3), 309–315.
23. Wicks JH (1966) Taxpayer compliance costs from personal income taxation. *Iowa Bus. Digest* (August), 16–21.
24. Wicks JH (1965) Taxpayer compliance costs from Montana personal income tax. *Montano Bus. Q.* (Fall), 36–42.
25. Yocum JC (1961) Retailers' costs of sales tax collection in Ohio. Ohio University Bureau of Business Research, Columbus.

ADDRESS FOR CORRESPONDENCE: *Michael Godwin Esq, Centre for Fiscal Studies, University of Bath, Claverton Down, Bath BA2 7AY, UK.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 399–406
© Pergamon Press Ltd. 1978. Printed in Great Britain

0305-0483/78/1001-0399$02.00/0

Measurement of Changes: What Criteria Do We Adopt?

EN CORLETT

AT PARSONS

University of Birmingham, UK

(*Received April* 1978)

Many changes in industrial working methods are now under discussion which are not easily evaluated and management is concerned about the probable return from investing in these methods. It is proposed that, whilst it is clear that improving work performance is a benefit, it is also of benefit to achieve changes in attitudes which correlate more closely with the objectives and state of the enterprise. Increased agreement in this respect is stated to lead to increased stability for the production process as there is less likelihood of disruption of output due to inappropriate decisions and actions. Since the loss due to various forms of interruption in output can be assessed it is described how the benefit of attitude change might thus be measured. A model for the relationships relevant to ergonomic and behavioural science changes in industry is given and its potential is illustrated by examples.

INTRODUCTION

NOT SO very long ago it was managerially acceptable to assess the costs of changes in work by simple costings along one-dimensional criteria. Perhaps the most popular method for evaluating the potential of a change was to assess the change in cycle time or estimated output arising from the proposal, using one of several estimating systems, calculate the direct financial return accruing from this and compare it with the current one. If the return exceeded the direct costs of change by an acceptable margin, the change was instituted.

If the route to improved performance was only through a step change in cycle time, which could be observed as changes in an operator's actions, then this procedure was a relevant one, although there is room for discussion regarding its implementation. Even if it is not a step change, but a gradual change through learning, providing there is only one dimension to the problem, the estimation of future gains is in some cases possible; see, for example, de Jong [3]. Our experience, and the literature, tell us that this direction for improvement in performance is not the only one, however. Too often do we find that there is a large gap between potential and actual performance, the reduction of which has no connection with cycle time. Equipment utilisation is a well known example; it is not unusual to find 60 or 70% utilisation in machine shops [1] and 15% has been demonstrated [4]. The improvement needed could arise, for example, from better scheduling, reorganisation of the payment system, changes in the distribution of setting and maintenance activities between operators and service personnel, but these changes would not noticeably affect the operator's rated performance.

In predicting the effects of measures to improve *utilisation*, considerations of movement times are not usually helpful. In many situations an operator will, by practice, achieve cycle times which will not significantly be bettered by work place modifications, even though the present workplace is demonstrably inadequate ergonomically. The adaptability of the human body is considerable and there is no innate mechanism which prevents people injuring themselves from undesirable activities. An

interesting example of the difficulty of assessing the payoff from ergonomic changes arose in a press investigation. To modify an 80 ton C-frame press to improve operator–machine relationships was going to cost £1,000. The company concerned asked their MTM analysts to evaluate the payoff, which came to £60 per year so they did not implement the changes. Later, in another company, a similar machine was changed to include the desirable characteristics, giving an improved output of about 10%, although without any noticeable change of cycle time.

The example quoted illustrates a question which is asked frequently in the work study field—will improved performance arise as a result of certain changes? The decision is often taken as a result of using certain techniques, chosen more for their availability than their appropriateness. The technique defines, implicitly, the model of the process, which often is one-dimensional. What is perhaps of greater importance, however, is that the lack of appropriate techniques causes large areas of the problem to be ignored, or given a cursory qualitative assessment. The decision is taken only on what can be measured.

Problems in industry are not simple cause and effect relationships—a fact which is widely appreciated—and we are faced with serious demands that they should not be treated as such. Management wants to know the real and total effects, for example, of ergonomic changes, of modifications to work organisation, whether certain alterations lead to safer work or less errors, and the results of good housekeeping. All these are aspects of work which affect human performance and those functions which are concerned with the design, organisation or management of people in industry must equip themselves to handle these problems.

TABLE 1. SOME OUTCOMES DESIRED BY MANAGEMENT OR OPERATORS AS A RESULT OF WORKING CHANGES

Factor	Management	Operators
Reduced errors	x	x
More output in a given time	x	x
Reduced effort		x
Improved safety	x	x
Reduced job stress		x
More interesting work		x
Reduced turnover costs	x	
Reduced training	x	Sometimes
Reduced interference with output	x	

Note: Reduced effort is a common work objective, but the criterion is increased output; it is not an end in itself.

These demands for better measures will come from others than management; they come also from the users of these working situations. We may note, in Table 1, that the current situation reveals some discrepancies between what management is likely to be interested in measuring, and what the shop floor may see as relevant. The industrial trend is slowly but surely to draw these two viewpoints together. The effect will be to call for a wider availability of skills in the personnel engaged in the evaluation and development of human industrial performance.

COST BENEFIT MODELS

The continuity of a profitable business can be seen as a target, but measuring profitability does not help us to decide whether to tidy the factory or change a workplace design. Neither does total output or hourly output help—they are too remote from the changes and may well have been more affected by events other than those which were of concern. If such changes are made we require some direct measures to assess them, used in relation to a framework which relates them to overall objectives and provides opportunity for them to be evaluated in that framework if required. In the case of the automobile engineer, interested in efficient high output from his engine, he monitors not only its BHP and fuel consumption, but also the temperature, exhaust gas condition and many other factors. From these he directly assesses particular aspects of its functioning, aspects otherwise closed to him if he has only data on BHP and fuel. What framework is there in the human performance area which will support a series of measures each connected to each by recognisable influences, and the whole being an operational analogue of the factory, which enables us to interpret and predict the overall performance by an analysis of relevant components?

This is the concept underlying the problem addressed in this paper. It can be separated into sequential stages. Firstly there is the problem of techniques with which to measure the appropriate criteria resulting from ergonomic changes. These criteria may be direct indicators of performance, or they may be factors recog-

Vol. 6
1978

nised as interacting to influence performance indirectly. For the latter it is then necessary to define the relationship to performance—how this influence is achieved—and it is not yet possible to provide other than qualitative associations between many of these factors. Figure 1, from Parsons [7] illustrates these associations.

The second problem is that of putting a value on the costs of the first stage and the benefits achieved. This would appear to need a quantification of the links between the various factors of Fig. 1 so that a given change in a causal variable or a causal intervening variable can have a defined influence on the cost variables.

In the study from which Fig. 1 is drawn [Parsons, 7], it was concluded that the current state of ergonomics technology and that of other human sciences did not permit this to be done. Not only were there insufficient techniques but there was also very little knowledge of the relative loadings between the various items which contribute to subsequent factors in the sequence.

Thus the apportioning of costs, and the resultant opportunities that this would provide to decide on directions for change, is not yet possible, at least in sufficient generality to make a standardised technique practicable.

However, this does not mean that no useful analysis is possible, for there is another direction which can be taken. Can it be shown that there are monetary benefits, even though all of them might not be specifically calculable? If they are not able to be expressed in cash terms, they cannot form part of a profit and loss statement, but can they be shown to contribute to these other major objectives of a company, its continuity and growth?

We can start by assuming that continuity and growth will be achieved most readily where the objectives of all those who maintain that continuity are closely in accord. This is applicable both in the physical and social spheres. The most ergonomic workplace permits maximum effort to be devoted to the job in hand rather than overcoming the inefficiencies of the equipment. Equally where the objectives of the work force are most in harmony with the objectives of management, less divergence in direction is probable. (Note that this last point does not call for an identity of views, but an acceptance of common objectives, or

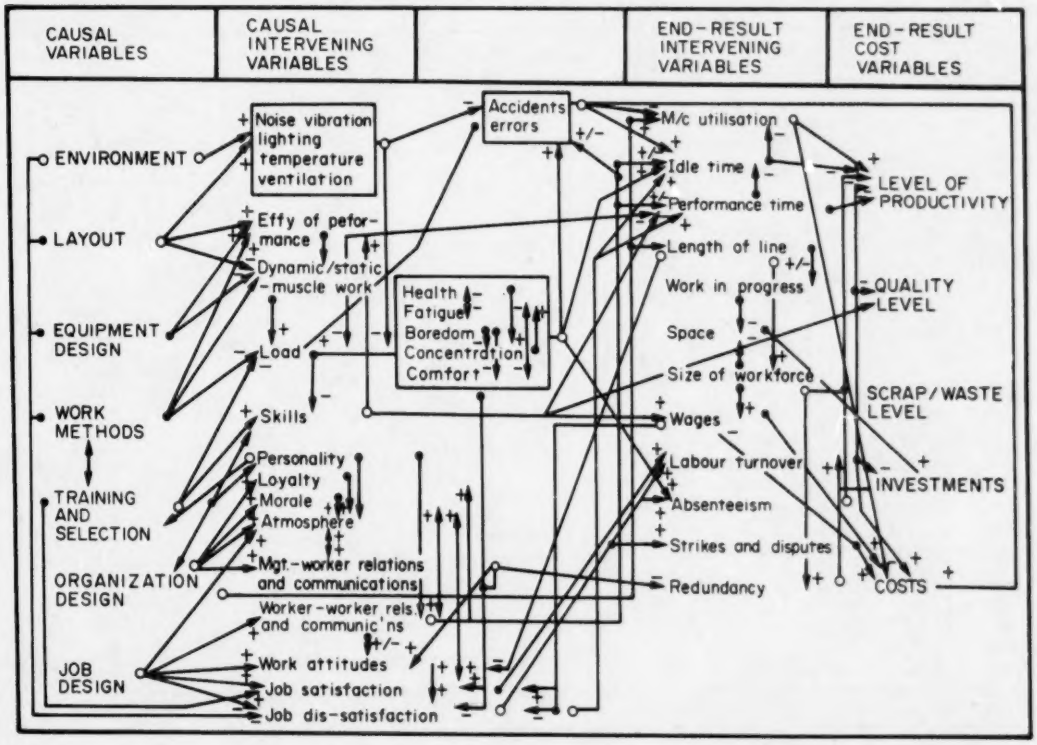


FIG. 1.

of objectives which result in a commonalty of achievement.)

The concept of accordance can thus be called upon to help us to evaluate ergonomic changes. To quote from Parsons [7, pp. 42–43]: "Accordance, in this context, is concerned with ability and willingness to adapt and to match performance and attitudes to the requirements of a situation, and thus two separate components of accordance can be identified. One, a performance–accordance factor, refers to aspects of an individual's behavioural performance–his level of skill, his response to high noise levels–and therefore also to the ergonomic factors which affect his performance, environment conditions, workplace layout, equipment design, training and selection. Secondly, an attitudinal–accordance factor which reflects the individual's subjective responses to all aspects of his work situation and acts as a stimulus in that it affects his performance within the work situation. All aspects of ergonomics affect this factor and it is the individual's perceptions of his environment, in the broadest sense, which determine his attitudes towards that environment and thus his responses, subjective and behavioural".

Using these concepts, it can be said that if the accordance level, for one or both factors, is raised then the change has benefited the firm. The change in level may result in a behavioural change, or changes, in some factors which can be measured on scales interpretable in cost terms; if so, this is additional evidence. But it is persuasive evidence if it can be shown that the proposed changes affect a number of factors which improve the level of accord, either to result directly in improved output, or to present to management and workforce increased opportunities for harmonious cooperation towards the company goals.

It is clear that these two accordance factors are not two independent factors which arrive at static levels, but that they are dynamic and inter-acting, having effects not only on each other but via loops around the system, as in Fig. 2. This is a simplified, and slightly modified version of the earlier figure where the links, though still existing, have not been drawn in.

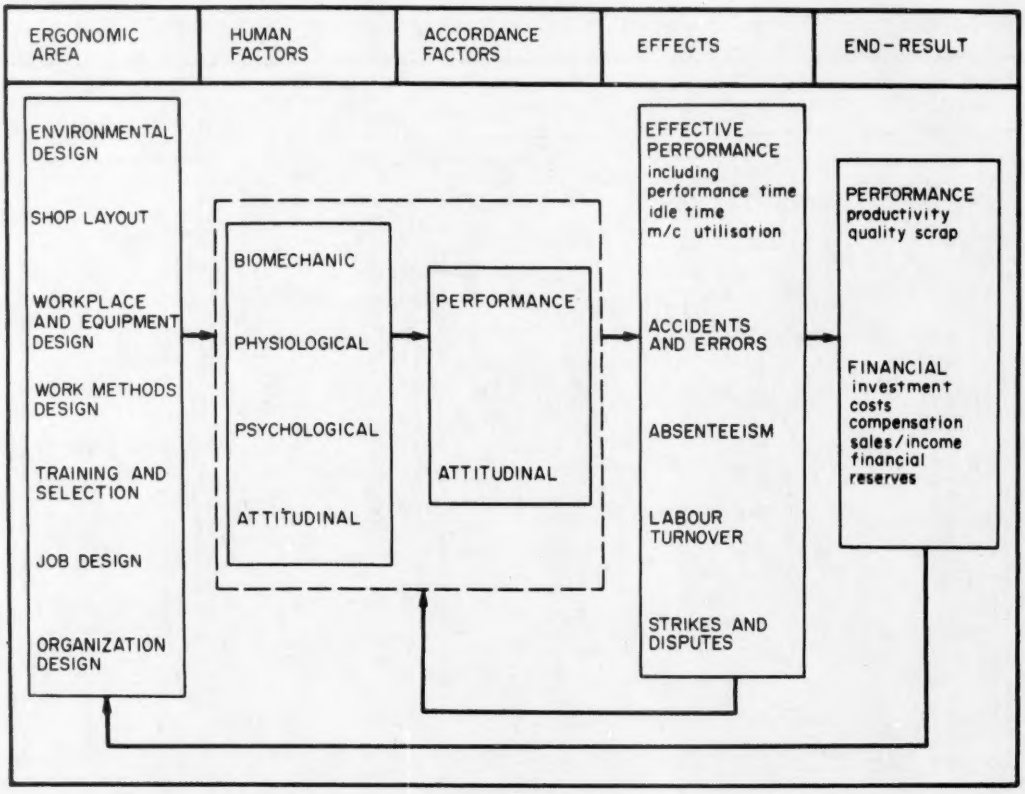


FIG. 2.

Let us now discuss these two factors.

Performance–accordance

Before discussing how performance-accordance might be assessed, it is useful to recall our objectives. We wish to introduce changes into the current situation which will permit improved performance as a result of a better match between the equipment, physical environment and methods, and the person doing the job. This requires us, initially, to deduce which out of the many available factors will be used, but we also wish to estimate the numbers of factors which will be influenced to improve this accord. We may find that optimum performance is associated more with factors not amenable to numerical description than with the one or two we can add up.

The assessment of performance–accordance incorporates areas of work relatively familiar to ergonomists and work study specialists. It is here that many of the techniques of performance and man-machine analysis are appropriate. (These do not require enlargement, a review of some recent techniques will be found in [8].) However, the usual techniques are those which provide numerical values, research and practice being hungry for numbers, and it has already been pointed out that there is a wider field to be covered. Thus the chronocyclograph, normally used to provide movement time and velocity data, when analysed particularly for body movements, gives us evidence of equipment accessibility. Posture itself is a factor which adds to work load, and inadequate postures have adverse long term effects. Short term studies may be inadequate to reveal these effects and an epidemiological approach may be called for. The results of studies of these factors are of importance for performance–accordance since they have a wide spreading influence, as Fig. 1 demonstrates. Early evidence of an inadequate biomechanic fit, which gives rise to bad postures, may sometimes be revealed by measures of industrial comfort (as described by Corlett and Bishop, 1976). This technique uses an arbitrary comfort scale to examine pain levels and pain distributions as they change through the work period. Meaningful comparisons between original and modified situations can be made, and correlated with changes in output via increased machine utilisation. As with some of the other factors just mentioned, this could be expected to interact with attitudinal–accordance; nevertheless, it has a strong direct influence on performance–accordance.

Attitudinal–accordance

Attitudes are seen as affecting behaviour, and consequently the physical performance of the factory system, as well as how the individual approaches and responds to all events in that system. A maximum level of attitudinal–accordance is taken to be that which enables some optimum level of performance to be achieved and maintained. It is begging the question to move on to job satisfaction and dissatisfaction as indicators of accordance, since these authors would not accept that they were necessarily correlated with performance. These two areas will be the major considerations in this section, however, because some breadth of understanding is developing regarding how satisfaction is influenced, as well as its relations with performance, and the discussion can stand as a model for considering other factors which are involved in attitudinal accordance.

At the moment it may be unwise to accept some current beliefs on what factors influence job satisfaction; the present emphasis on job content and enrichment is welcome but perhaps too single-minded. Since we are, as with performance–accordance, concerned to achieve a better match between worker and management objectives, we require answers from questionnaires (or other instruments) which relate the attitudes and specific aspects of the work situation wherever possible. At the same time, a general estimate, on a satisfaction–dissatisfaction continuum, is necessary, since it is not possible to add up the scores for the separate items on a questionnaire and call this the satisfaction level. This is not just because items which provide satisfaction may not be the same as those which dissatisfy, as Herzberg [5] and others have demonstrated. All items contribute to the overall level, but we do not know if we have them all on our list; neither do we know the weightings to apply to those we have.

The results from such a study indicate where satisfaction or dissatisfaction is expressed and the factors with which each is associated. By using Fig. 1 (and eventually a more appropriate Fig. 1 developed from the results of

studies) the probable interactions resulting from changes in the measured factors may be deduced and those dimensions available for change can be recognised. The likely effects of introducing changes on the direction in which attitudinal–accordance would be influenced can also be estimated.

The measures in this area include finding out what the objectives are of people when at work, and what their perceptions of work are. Such data indicate whether the objectives of company and workers (at all levels) are matched and permit changes to be instituted so that striving for company and for personal goals are not opposing activities, but as closely in harmony as is practical. Motivation to work is important and relevant; so too is the ability to achieve, and at this point it must be clear that increasing motivation without providing increased opportunity to achieve is creating the scenario for frustration. Increased opportunity to achieve does not mean just a larger pile of available components, vital though the supply of work is in this context; it means, amongst other things, reducing the interference between the worker and the finished component. If the classical ergonomic aspects of workplace design have been attended to, if the organisation has been changed to provide the worker with the opportunity to reduce his own delays on set-up, breakdown or material supply and if knowledge of performance in quality as well as quantity terms is quickly available for him to do something about it, then increased motivation has something to work on. If, additionally, the changes are such that the increased performance does not lead eventually to decreased employment, the usual result perceived by workers as the outcome of meeting management's objectives, then there is some chance that changes in attitude–accordance are of value. The industrial climate has been changed to be more in accord with that required for effective performance.

TWO EXAMPLES

An examination of two studies in which costs and benefits have been assessed may help to show the directions in which the concepts of this paper can be developed.

Costing the benefits of training

This study, by Jones and Moxham [6] was a major cost–benefit analysis of training.

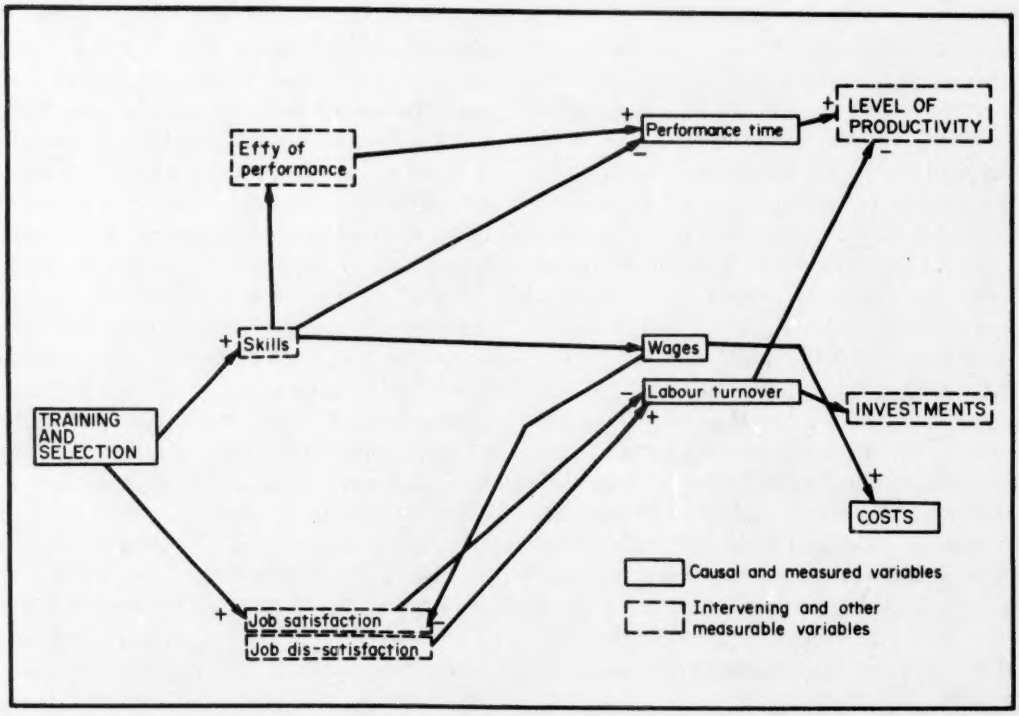


FIG. 3. *Costing the benefits of training (Jones and Moxham, 1969).*

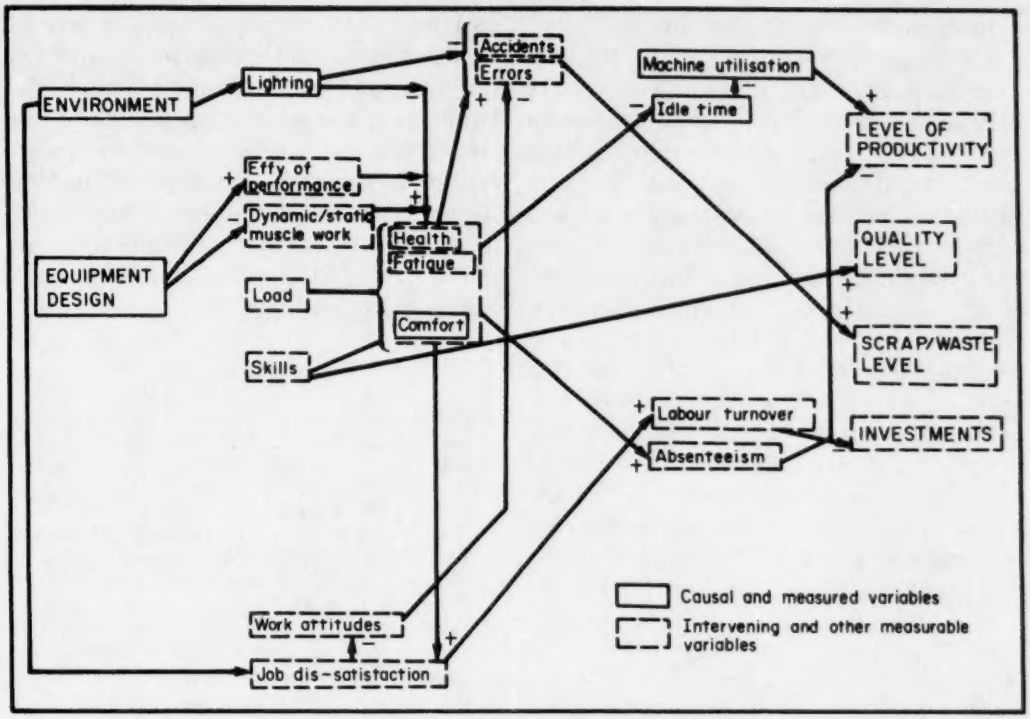


FIG. 4. *Ergonomic analysis of a spot welder (Bishop, 1973).*

Figure 3 shows the factors which were measured, both the dependent and independent variables, as well as the major links between these factors. It will be seen that the benefit was demonstrated mainly in terms of performance time and labour turnover. This latter is influenced via job satisfaction, which might have been measured directly, particularly as it would probably respond more quickly to changes in training than would turnover. Productivity also might have proved a useful additional measure, as it would have incorporated the total usage of the equipment and personnel rather than the partial measure of performance time. It will be realised that this latter method does not bring into the analysis either machine usage or the work/rest ratio of the operator. The total benefit to the company is probably conservatively estimated by the measures of Jones and Moxham [6] (no bad thing of course in the present state of cost–benefit analysis).

The diagram of Fig. 3 illustrates hypothesised links between the measured factors, important output variables and some intervening variables. These last may not be all that are involved, but only a study of an actual situation will reveal if there are others of importance.

If a more detailed cost–benefit analysis is required, the diagram demonstrates where techniques will have to be developed. The influence and contribution of some factors are clearly to improve performance, by matching better the worker and the job. Other factors modify attitudes to work, a no less important change in many situations, yet the costing of this is relatively crude, as the diagram suggests. A recent investigation by Watts (1974) into the costs of labour turnover attempted to specify more precisely the factors influencing costs as a result of turnover, and ways of assessing acceptable levels of cost when attempting to reduce turnover.

Ergonomic analysis of a spot welder

When undertaking this study, Bishop [1] set out to provide some measures of benefit and cost. Attitudes were relatively less well explored than performance. A major aspect of the job which was much objected to was the physical pain arising from working the machines, a recognised influence on output as well as on departure from the job either tem-

porarily or permanently. The comfort "level" was measured, but its contribution to the total excess costs in the situation would need more detailed measures. The increased load from high discomfort levels plausibly degrades skill, giving rise to poorer quality welds. It has a widespread influence on productivity through attitudes to the work and the job. At each point, research might produce a measure either of the discrepancies between performance capacity and output, or between a balance of personal views which encourage cooperative working towards some common objectives and the balance currently held by the employees concerned. This latter assessment, a statement of attitudinal accordance, would be a step towards demonstrating the contribution of such changes to the long term performance characteristics of the working group involved.

CONCLUSIONS

This paper presents a direction for developing a cost–benefit analysis procedure. Particular cases can use the appropriate groupings of factors from Fig. 1, modified as experience develops. Changes in the levels of these factors may be assessed quantitatively in some cases but along nominal or ordinal scales in others, and the directions of change identified. Performance measures can be supported by indications of attitude changes, which are seen as enabling the performance changes to be more readily achieved and maintained. It is argued that although attitudes are not yet measurable directly in cost terms they have a value, or cost, for the firm and it is of value to it to study and record how they are influenced. Indeed it is more relevant and indicative of total benefit to evaluate the results of ergonomic changes by this method than by the currently available costing procedures.

REFERENCES

1. Bishop RP (1973) The ergonomics of foot operated spot welders. Unpublished *MSc. Report*, Dept of Engineering Production, University of Birmingham, UK.
2. Corlett EN & Bishop RP (1976) A technique for assessing postural discomfort. *Ergonomics* **19**, 175–182.
3. De Jong JR (1957) The effects of increasing skill on cycle time and its consequences for time standards. *Ergonomics* **1**, 51–60.
4. Dudley NA (1970) Comparative productivity analysis. Study in the UK, West Midlands engineering and metalworking industries. *Int. J. Prod. Res.* **8**, 397–403.
5. Herzberg F (1968) One more time: how do you motivate employees? *Harv. Bus. Rev.* **46**, 53–62.
6. Jones A & Moxham J (1969) Costing the benefits of training. *Personnel Mgmt* **1(4)**, 22–26.
7. Parsons AT (1973) Ergonomic cost benefit analysis. Unpublished *MSc. Report*, Dept of Engineering Production, University of Birmingham, UK.
8. Singleton WT, Fox JG & Whitfield P (Eds) (1973) *The Measurement of Man at Work*. Taylor & Francis, London.

ADDRESS FOR CORRESPONDENCE: *Professor EN Corlett, Dept of Engineering Production, The University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 407–417
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0407$02.00/0

Structural Mapping Applied to Single Value Policies

PATRICK RIVETT

University of Sussex, UK

(Received November 1977; in revised form April 1978)

The paper is a development of the use of multi dimensional scaling techniques for multiple criteria decision. It takes the special case of policies to each of which a single value can be applied and assigns indifferences between all pairs of these policies based on a series of probability laws. It is shown that the mapping performs well not only in throwing to opposite ends the high and low value policies but also shows that the position of policy points along the principal axis of the map is related to the single values.

1. INTRODUCTION

THE METHOD of multiple dimensional scaling originally developed by Shepard [10] and Kruskal [6, 7] is widely used in psychological testing. Kendall [2, 3, 4, 5] has shown how this method can be applied to draw spatial maps, for example, to place in time order prehistoric graves according to the evidence they contain and also to construct two dimensional maps of the layout of mediaeval manors on the basis of fragmentary information. Rivett [8] has shown how Kendall's development can be applied to the selection of policies where there are multiple criteria of choice and this method has been developed in further work by Clarke and Rivett [1] and by an example in Rivett [9]. It is also inferred that the principal axis of the map corresponds in some way to the decision maker's axis of preference.

These examples have demonstrated the feasibility of constructing maps of points representing policies, in which the most preferred and least preferred policy points are thrown to the polar extremities. The input is a set of statements of those pairs of policies, the results of which the decision maker feels equally attractive (i.e. a concept of 'indifference'). It is not necessary that these indifference statements be transitive. In fact the method proposed will break down if there is complete transitivity. The robustness investigations detailed in Clarke and Rivett [1], show, in particular, that the maps remain fairly constant provided that each policy has an 'indifference' link assigned to at least two other policies.

2. OBJECT OF THE PRESENT STUDY

Consider a set of m policies, P_1, P_2, $P_3, \ldots P_m$, each of which can be assessed against n criteria, $C_1, C_2, \ldots C_n$. We can state the performance of these policies against each of the criteria in matrix form, in which each row represents one of the policies and each of the n columns represent the degrees of attainment (a_{ij}) of a particular criterion by the policies in turn.
That is:

		Criteria			
		C_1	C_2	...	C_n
	P_1	a_{11}	a_{12}	...	a_{1n}
Policies	P_2	a_{21}	a_{22}	...	a_{2n}
		⋮	⋮		⋮
	P_n	a_{m1}	a_{m2}	...	a_{mn}

It can be shown that if a decision maker puts the n policies in any rank order of preference

then there is a real transformation of the form

$$A_i = \sum_{j=1}^{n} f(a_{ij})a_{ij}$$

such that if policy P_1 is preferred to policy P_2 is preferred to policy P_3 etc. then $A_1 > A_2 > A_3 \ldots > A_n$

Proof

(i) If $n \geqslant m$, then consider

$$\begin{aligned} b_{11}a_{11} + b_{12}a_{12} + \ldots \quad b_{1n}a_{1n} &= A_1 \\ \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots \\ b_{m1}a_{m1} + b_{m2}a_{m2} + \ldots + b_{mn}a_{mn} &= A_m \end{aligned}$$

With the usual constraint on the rank of $\{b_{ij}\}$, if $m = n$, this is uniquely solvable; if $n > m$ there is an infinity of solutions for (b_{ij}).

(ii) If $n < m$, Let $m - n = p - 1$

Then consider

$$\begin{aligned} (b^1_{11}a_{11} + b^1_{12}a^2_{12} + \ldots b^1_{1p}a^p_{1p}) \quad + b_{12}a_{12} + \ldots + b_{1n}a_{1n} &= A_1 \\ \vdots \qquad\qquad\qquad\qquad\qquad \vdots \qquad\quad \vdots \\ (b^1_{m1}a_{m1} + b^1_{m2}a^2_{m2} + b^1_{mp}a^p_{mp}) + b_{m2}a_{m2} + \ldots \ldots \ldots + b_{mn}a_{mn} &= A_m \end{aligned}$$

(note, the same constraint on the matrix of the coefficients applies.

These two sets of (b_{ij}) give any required set of (A_i) which more than satisfies a simple rank ordering requirement.

This result is widely quoted but the author of this paper is unable to attribute an original reference to it. However, the significance of the result is to make one enquire what goes on in the mind of a decision maker when he ranks multiple criteria choices. Do we each have tucked away in our heads a transformation function which unconsciously applies itself to each of the choices? Does ranking, in fact, imply a single value? Unfortunately the literature of psychology seems unclear on this point, but in the light of the mapping research already mentioned, it seemed of interest to enquire what would be the shape of maps drawn when policies, in fact, have a single value assigned to them.

At the risk of appearing philosophical one might reflect at this point on the consequences of the result mentioned above. Cost Benefit analysis has been criticised in Rivett [8] as forcing measures on concepts which are, by hypothesis, incommensurable. There is an obvious paradox in the requirement of multiple criteria analysis that the criteria are incommensurable and the practical necessity of still making a choice. Perhaps the consequence of this result is that there is no such thing as a multiple criteria decision? Within our minds there is a transformation that converts differing amounts of chalk and cheese into a common scale. Hence the criticism of cost benefit analysis made previously by the author on axiomatic grounds is invalid if the hypothesis of incommensurability is untenable.

3. THE EXPERIMENT

Consequently we have taken 3 sets of 25 'policies'. We have assigned a single value to each policy and these assigned values lie between 0 and 100. In one set the values are 25 random normal deviates with a mean of 50, and standard deviation of 25. In the second there are 25 random rectangular deviates over the range 0,100. It was remarked in Rivett [8] that there seems a tendency for the mapping to throw to the perimeter of the maps policies with fewer links. Since the policies of extreme preference are likely to have fewer links, *a priori*, one wishes to ensure that the structural mapping process that has been proposed is not merely an expensive way of sorting out policies with few links. Hence a third set of policies has been studied in which the distribution of value is random 'U' shaped. This will have highly linked policies at each extreme and one can then test whether in these circumstances

TABLE 1. POLICY VALUES FOR EXPERIMENT

Policy		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Rectangular	(R)	0	3	12	16	18	22	23	26	27	29	31	33	37	44	52	55	56	57	63	68	70	74	84	97	99
Normal	(N)	0	7	23	27	28	29	30	34	35	37	38	39	42	46	51	53	54	55	57	60	62	65	76	98	99
U Shaped	(U)	1	4	7	10	13	15	19	24	29	35	42	50	58	65	71	77	81	86	90	92	93	95	96	98	99

TABLE 2.

Let the value of policy P_i be V_i. A link of indifference between two policies P_i and P_j will probabilistically depend on $|V_i - V_j| = V_{ij}$ (say). Let the probability laws be $P_1(V_{ij})$, $P_2(V_{ij})$ etc.

(1)	$P_1 = 0.5$,	$0 \leqslant V_{ij} \leqslant 25$;	$P_1 = 0 \quad V_{ij} > 25$
(2)	$P_2 = 0.75 - \frac{V_{ij}}{50}$,	$0 \leqslant V_{ij} \leqslant 25$;	$P_2 = 0 \quad V_{ij} > 25$
(3)	$P_3 = 1.0 - \frac{V_{ij}}{25}$,	$0 \leqslant V_{ij} \leqslant 25$;	$P_3 = 0 \quad V_{ij} > 25$
(4)	$P_4 = 1.0$,	$0 \leqslant V_{ij} \leqslant 6.25$;	
	$= 1.5 - \frac{2V_{ij}}{25}$,	$6.25 \leqslant V_{ij} \leqslant 18.75$;	
	$= 0$	$V_{ij} > 18.75$	
(5)	$P_5 = 1.0$	$0 \leqslant V_{ij} \leqslant 12.5$	
	$= 0$	$V_{ij} > 12.5$	

Figure 1 illustrates these laws rather more clearly.

policies with fewer links were forced to the centre of the map.

Table 1 shows the values of the 25 points in each of the 3 sets.

4. POLICY LINKS

The next step was to assign links between pairs of policies depending on different probabilistic criteria. Five probability laws were taken as shown in Table 2.

We, therefore, had three sets of points and to each was applied five different probabilistic laws resulting in pair indifferences being assigned. It will be noticed that law 1 is the most 'sloppy' and the assigning of links becomes more precise as one progresses through 2, 3, 4 to 5. The consequence is that in law 1 close points only stand a 50/50 chance of being linked whereas in law 5 close points are certain to be linked and distant points will never be linked.

It should be noted that, in establishing links by probabilities, on (rare) occasions nature had to be assisted. For example in case *N*, (Table 1) points 23 and 24 are value 'distant' 22 and a link had to be inserted in laws 4 & 5, as no link is possible at that distance under these laws.

5. THE MAPS

For each of these 15 sets of policy links a different random starting configuration was taken. Maps were plotted based on a local-global process as listed in Kendall's development of the MDSCAL algorithm with refinements by Professor Sibson of the University of Bath. A tertiary treatment of ties was used, and a series of 75 iterations minimising the Kruskal stress [6] was made. There were two points of interest. Firstly since the contention of Rivett [8] was that the principal axis of these maps should be the axis of preference, to what extent is position along this principal axis related to the preassigned value of the point in question? Secondly, to what extent does the mapping process throw to the polar extremities the points corresponding to extreme values? A set of some of these maps is now given in Figs. 2–17; the letter refers to the type of distribution and the number refers

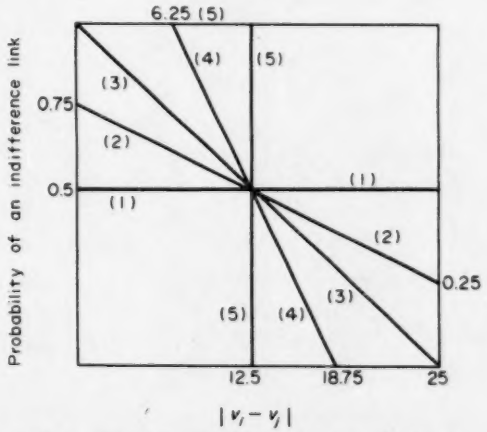


FIG. 1. *Probability Links between pairs of policies.*

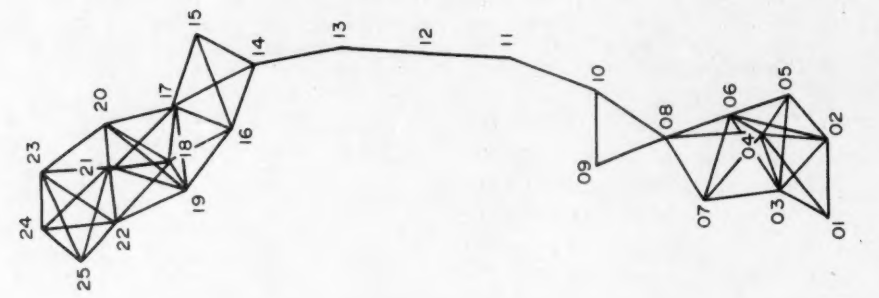


FIG. 5. *Value Distribution U. Probability Law (4).*

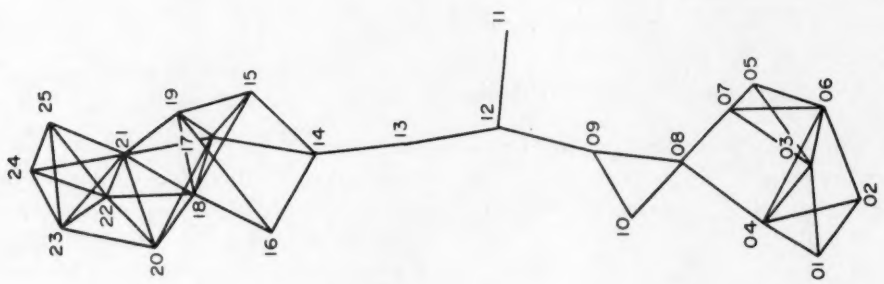


FIG. 4. *Value Distribution U. Probability Law (3).*

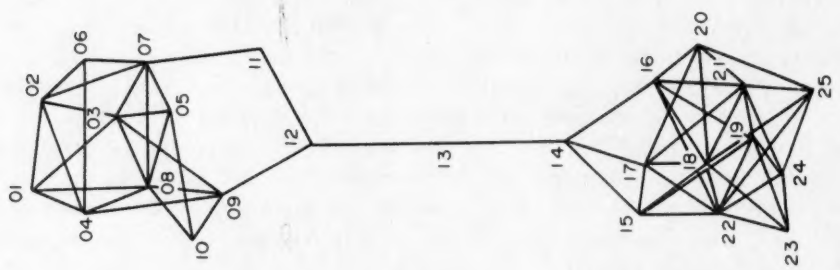


FIG. 3. *Value Distribution U. Probability Law (2).*

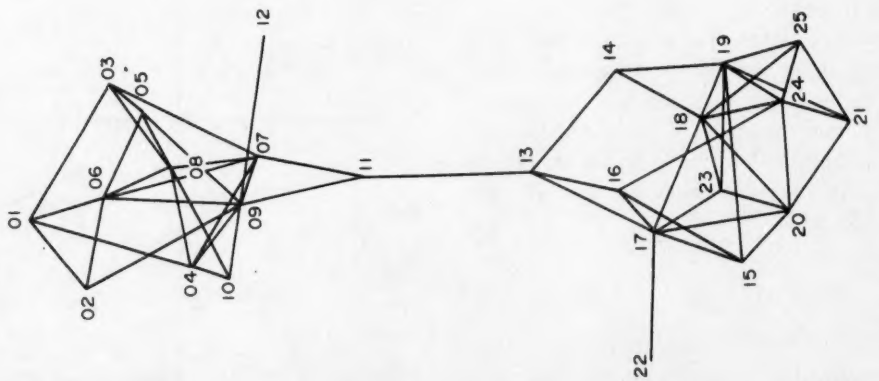


FIG. 2. *Value Distribution U. Probability Law (1).*

Vol. 6
1978

to the probability law (as above). In these figures, a line joining two points indicates that they were linked by the working of the probability law.

In order to test the first of these two properties, i.e. relation between vertical position (the maps are all drawn with the principal axis vertical) and pre-assigned value, correlation coefficients (Pearson's r) were calculated between the vertical distance up the map and the value. The absolute values of these 15 correlation coefficients are shown in Table 3.

It is interesting to look in diagram form at the linear relationship between the "correct" values of Table 1 and the values read off the principal axis, where vertical the scale is normalised to run from 1 to 99. Figure 17 gives one plot for each of the three distributions R, N, U.

The relation does look good, as well as having a high r. It is interesting to note that the form of the parent distribution is clear from the disposition along the principal axis. This is a rather remarkable property given the very weak form of the input data which are purely

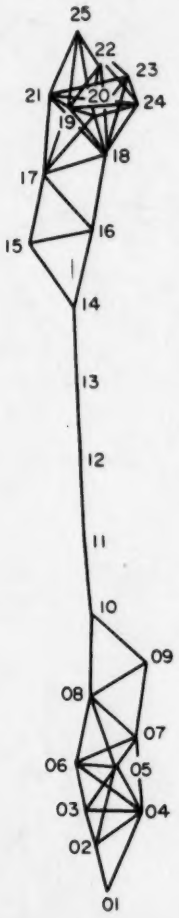


FIG. 6. *Value Distribution U. Probability Law* (5).

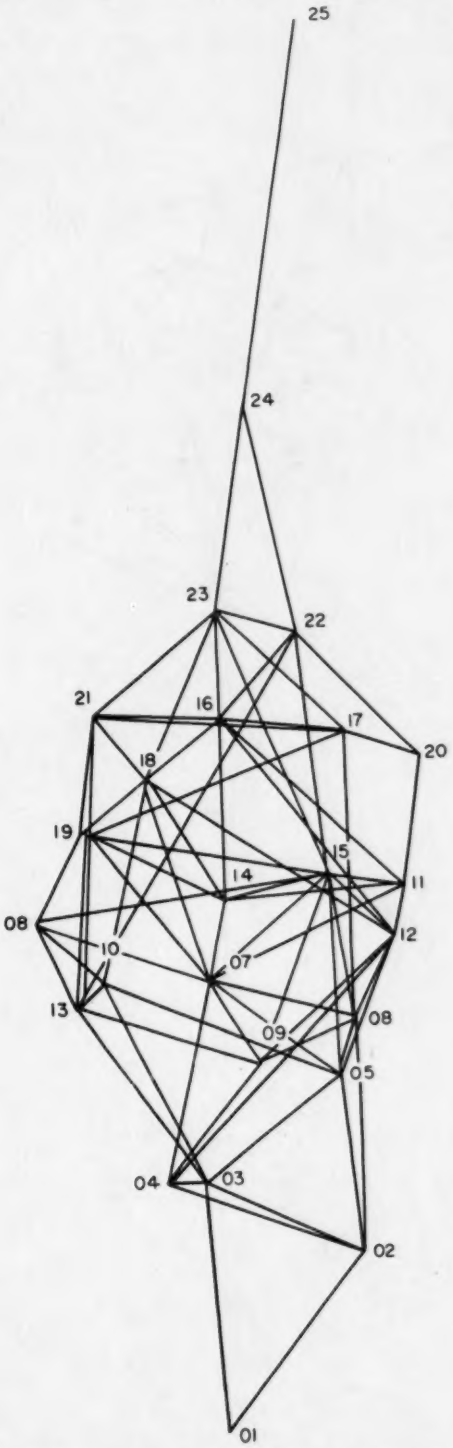


FIG. 7. *Value Distribution N. Probability Law* (1).

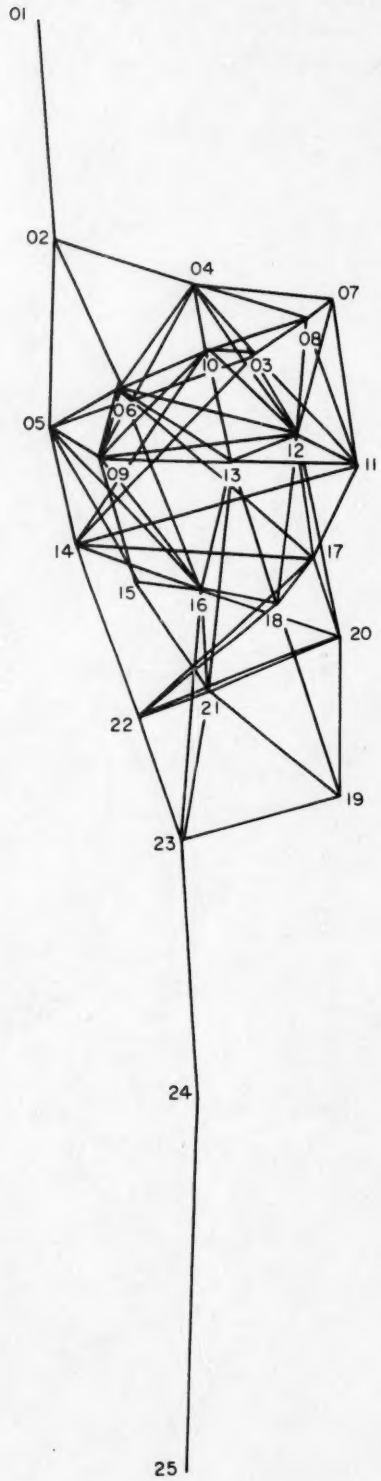


FIG. 8. *Value Distribution N. Probability Law* (2).

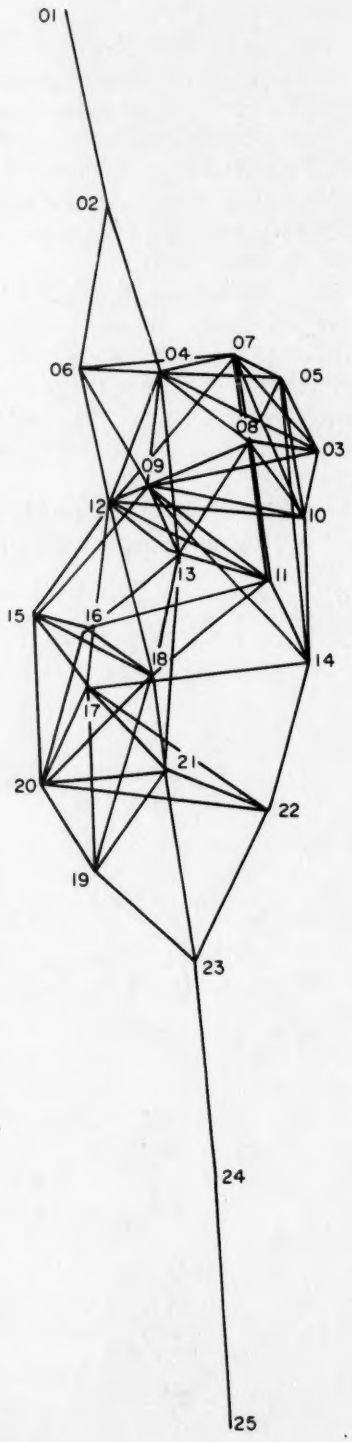


FIG. 9. *Value Distribution N. Probability Law* (3).

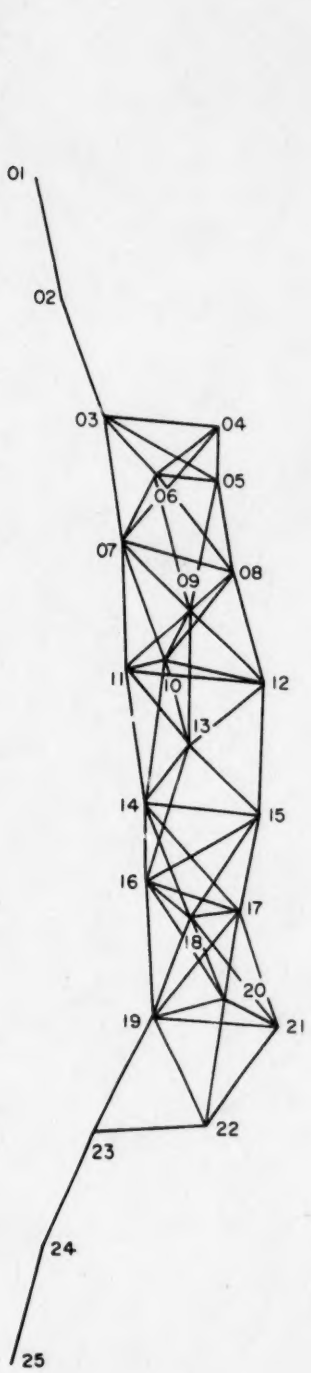


FIG. 10. *Value Distribution N. Probability Law* (4).

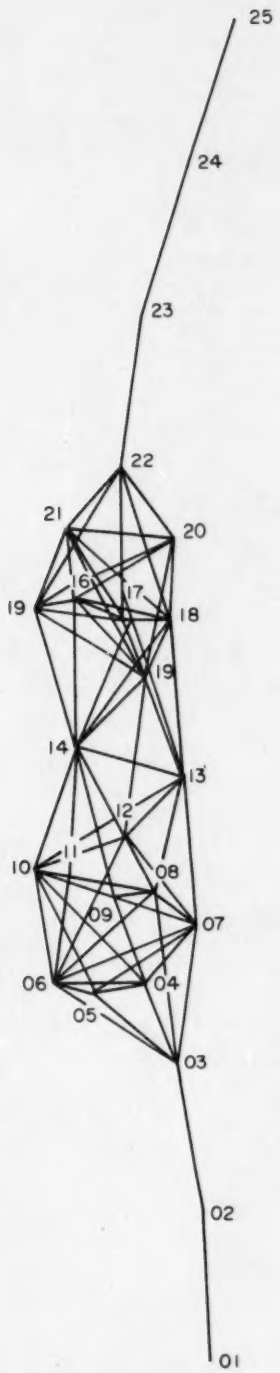


FIG. 11. *Value Distribution N. Probability Law* (5).

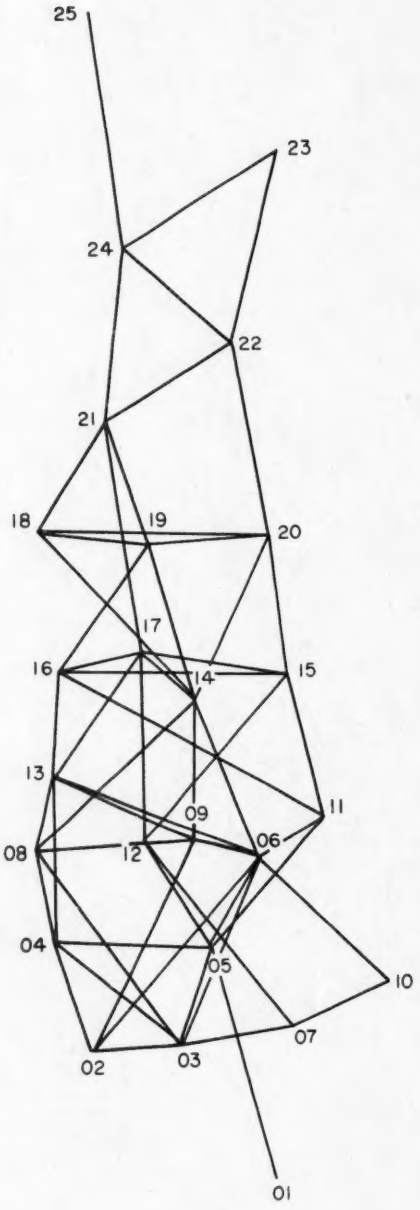


FIG. 12. *Value Distribution R. Probability Law* (1).

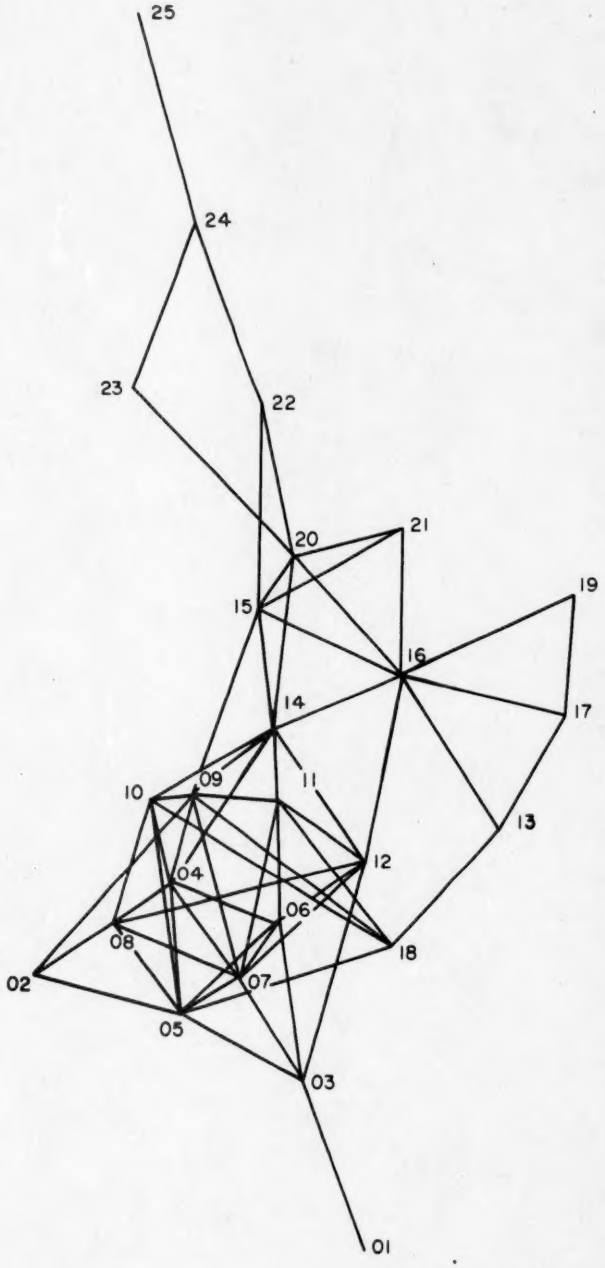


FIG. 13. *Value Distribution R. Probability Law* (2).

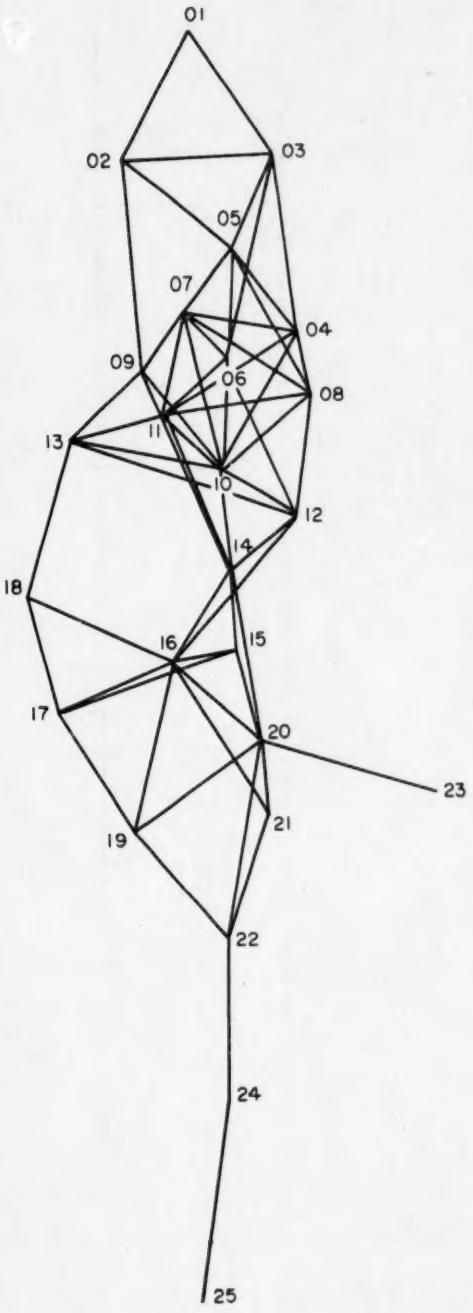


FIG. 14. *Value Distribution R. Probability Law* (3).

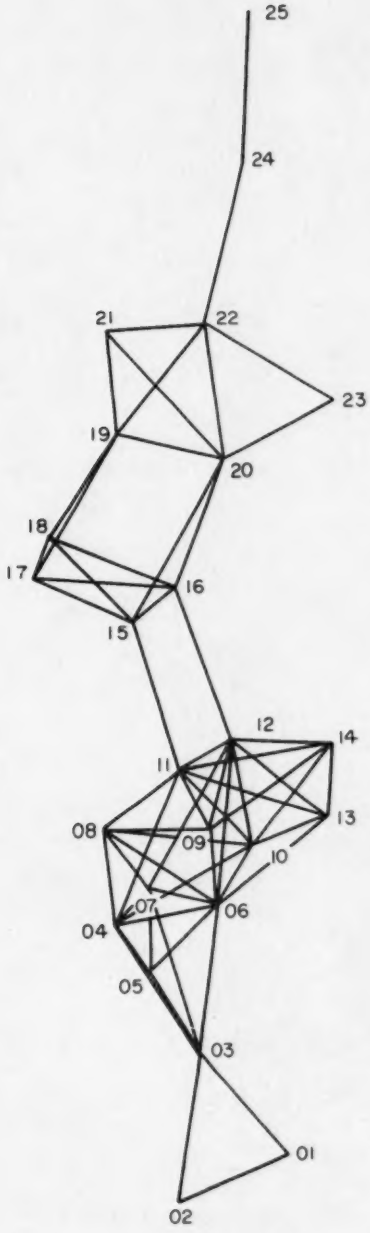


FIG. 15. *Value Distribution R. Probability Law* (4).

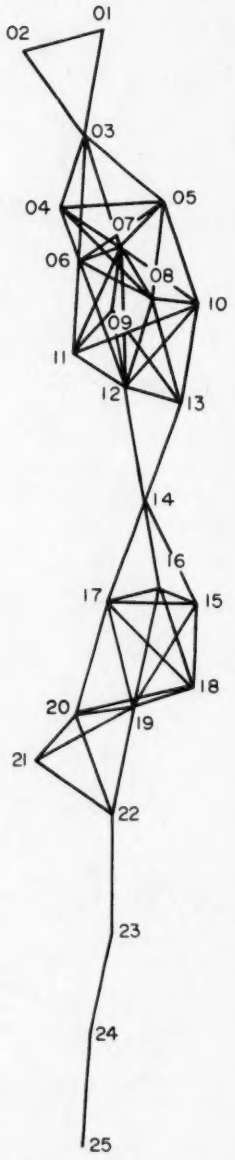


FIG. 16. *Value Distribution R. Probability Law* (5).

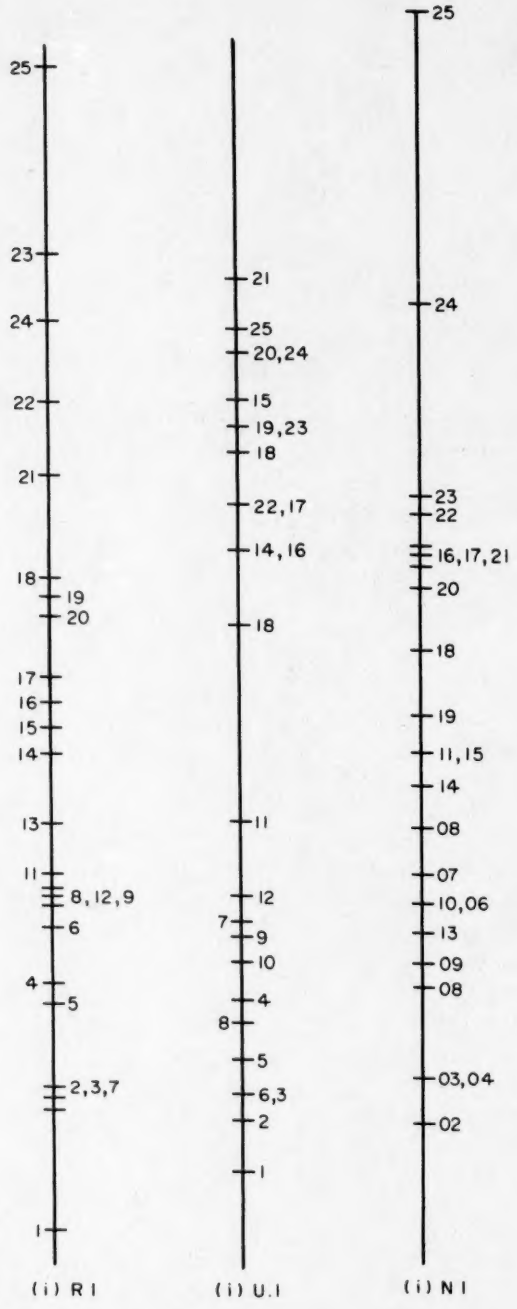


FIG. 17. *Vertical Distribution of Points in Figures* 1, 6, 11.

TABLE 3. PEARSON'S *R* FOR VALUE AGAINST VERTICAL POSITION

	Probability law				
Distribution	1	2	3	4	5
R	0.964	0.940	0.974	0.919	0.995
N	0.950	0.951	0.976	0.976	0.921
U	0.963	0.982	0.993	0.977	0.998

Vol. 6
1978

TABLE 4. EXTREME POINTS IN MAPS

	Probability law				
	1	2	3	4	5
R	1, 3, 2, 7 22, 24, 23, 25	1, 3, 5, 4, 22, 23, 24, 25	1, 3, 2, 5 23, 22, 24, 25	2, 1, 3, 5 21, 24, 25, 22	1, 2, 3, 5 22, 23, 24, 25
N	1, 2, 3, 4 22, 23, 24, 25	1, 2, 4, 7 19, 23, 24, 25	1, 2, 6, 7 19, 23, 24, 25	1, 2, 3, 4 22, 23, 24, 25	1, 2, 3, 5 22, 23, 24, 25
U	1, 2, 6, 2 24, 20, 25, 21	1, 6, 5, 2 22, 23, 24, 25	2, 1, 4, 6 22, 23, 24, 25	2, 6, 3, 1 25, 22, 18, 20	1, 2, 4, 3 21, 23, 22, 25

qualitative; i.e. the parent distribution of value can perhaps be deduced from information on *pairs*.

The second of these properties was tested by taking the four extreme "North" and "South" points in each of the maps. For the mapping process to be precise these should respectively be points 1–4 inclusive (in that order) and 22–25 inclusive (in that order). What actually happened is shown in Table 4.

6. CONCLUSIONS

Looking at the 15 maps it is obvious that one can tell from any one map the form of the parent distribution. It does seem quite remarkable that mere statements of non-transitive indifference form, (which logically are the weakest statements one can make) should lead to such a precise and robust conclusion. Secondly the correlation coefficients are all extremely high, so high that there is almost no need to test their significance (they all have significance levels of at least 0.97). Although the selection of the polar quartets is not always the precise quartet one would wish, nevertheless if the extreme points were taken in each diagram the results, and corresponding errors, would be as follows. (Note, by error we mean that if the value of the correct extreme point is x and the value of the extreme point in the diagram is y, then the error is $|lx - yl|$).

EXTREMES AND ERRORS IN DIAGRAMS

Run	Lower extreme	Error	Upper extreme	Error
*R*1	1	0	25	0
*R*2	1	0	25	0
*R*3	1	0	25	0
*R*4	2	3	22	25
*R*5	1	0	25	0
*N*1	1	0	25	0
*N*2	1	0	25	0
*N*3	1	0	25	0
*N*4	1	0	25	0
*N*5	1	0	25	0
*U*1	1	0	21	6
*U*2	1	0	25	0
*U*3	2	4	25	0
*U*4	2	4	20	7
*U*5	1	0	25	0

It must be remembered that the probability laws are rather 'sloppy'. The total range of the data is 100 and a point may be linked to other points with values of 25 in either direction, that is *half* the total range. It is suggested that the mapping process does look encouraging under these rather severe conditions.

Acknowledgements—I would like to thank David Clarke of Sussex University for his assistance with the computer runs and Professor DG Kendall of Cambridge University for making available his computer programmes.

REFERENCES

1. CLARKE D & RIVETT BHP (1977) A structural mapping approach to complex decision making. *Op. Res. Q.* **28**(4) (i).
2. KENDALL DG (1975) The recovery of structure from fragmentary information. *Phil. Trans. R. Soc. A.* **279**, 547–582.
3. KENDALL DG (1976) Tagungsbericht über Mathematische Stockastik. *Mathematisches Forschungsinstitut Oberwolfach*, pp. 6–7.
4. KENDALL DG (1976) Computer techniques and the archival map reconstruction of Mycenaean Messenia. To appear in *Proc. of the First Colloquium on Mycenaean Geography*.
5. KENDALL DG (1971) Construction of maps from 'odd bits of information'. *Nature* **231**, 158–159.
6. KRUSKAL JB (1964) Non-metric multidimensional scaling—a numerical method. *Psychometrika* **29**, 115–129.
7. KRUSKAL JB (1964) Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. *Psychometrika* **29**, 1–27.
8. RIVETT BHP (1977) Policy selection by structural mapping. *Proc. R. Soc. London A.* **354**, 407–423.
9. RIVETT BHP (1977) Multidimension scaling applied to the selection of policies where there are many objectives. *Omega* **5**(4), 367–379.
10. SHEPARD RN, ROMNEY AK & NERLOVE SB (Eds) (1972) *Multidimensional Scaling: Theory and Applications in the Behavioural Sciences*. Seminar Press, New York.

ADDRESS FOR CORRESPONDENCE: *Professor BHP Rivett, The University of Sussex, Operational Research, Mantell Building, Falmer, Brighton BN1 9RF, UK.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 419–426
© Pergamon Press Ltd. 1978. Printed in Great Britain

0305-0483/78/1001-0419$02.00/0

Visual Interactive Simulation An Aid to Decision Making

RD HURRION

RJR SECKER

University of Warwick, Coventry, UK

(*Received January 1978; in revised form April 1978*)

This paper describes a technique of visual interactive simulation. The technique allows a simulation model to be dynamically displayed on a visual display unit. The user can watch the model progress through time and has the ability to stop, interact and try alternative decision strategies. The paper shows that by watching a simulation model progress through time and having the ability to interact with it, a user can improve his analysis and understanding of the original problem situation.

INTRODUCTION

THE USE of simulation has long been established as a useful method of giving insight into complex problem situations. However, the technique of simulation does not usually allow a decision-maker/manager to see how such a complex problem is developing through time, nor does it give him the ability to interact with it. The technique only gives him statistical answers at the end of a set of particular simulation experiments [3, 4].

The approach when using simulation to analyse a complex situation would normally be:

(i) agreement of the problem situation with the decision-maker/manager/client;

(ii) development of simulation model;

(iii) experimentation;

(iv) report on the conclusions and recommendations to the decision-maker/manager.

This approach means that the decision-maker is not normally a part of the simulation development, and his experience and judgement cannot be used to directly assist the study. Thus, any conclusions obtained by the Operational Research Scientist must be taken on trust. If the conclusions do not agree with the intuition or practical judgement of the decision-maker a confidence gap will appear regarding the use of the model.

We are not suggesting that all Operational Research studies are carried out in isolation from the decision-maker. One hopes that this is far from true! However, the very nature of simulation studies means that a significant part of the analysis must appear as a 'black box'. For this reason many Operational Research simulation studies fail. The simulation model might run successfully, but overall, the study may not reach an implementation stage due to the decision maker/manager's lack of confidence in the model.

With these views in mind a visual interactive simulation system is proposed. The basic philosophy of the approach is that a decision-maker/manager will be able to watch a simulation model of his problem situation develop through time. This is achieved by using a visual display unit. The decision-maker can watch the progress of the model and can himself contribute to the validation of that model. Hence, he will have more confidence in its use, because of his own participation.

He will be able to use his knowledge and experience to interact with the model in order to explore alternative strategies, decisions or directives.

THE CURRENT PHILOSOPHY OF VISUAL INTERACTIVE SIMULATION

The philosophy behind the simulation system is to have:

(i) a simulation language in which it is possible to write complex industrial problems;

(ii) the ability of a one-to-one correspondence between elements in the model and elements as described visually on the visual display screen;

(iii) flexibility at run time.

The event scheduling, process interaction and activity scan simulation methods [1, 7] have all been used for interactive model design work but the activity scan approach gives an easy method of model building and has distinct advantages for interaction at run time. The activity scan approach for simulation work consists of testing activities to see if they can occur. If an activity can occur then the appropriate change to the model is made. The typical structure for an activity is shown in Fig. 1.

The activity tests contain the logic to ensure an activity can take place. If a one-to-one correspondence exists between elements in the model and elements on the visual display screen, then modification of activity tests is possible. This gives flexibility in the model structure at run time.

The one-to-one correspondence between the model and its visual representation is obtained by giving a unique string of characters to each entity in the model. Each set or queue in the model is also given a unique reference. Thus it is possible to describe an ordered set or queue on the screen by first showing the queue name followed by a list of the entities that appear in that queue.

The methodology behind the interactive approach is that the simulation model will run as determined by the current algorithms driving it (i.e. identical to a batch mode simulation). When the model changes with time, these changes are shown on the display screen by elements of particular queues or sets joining other sets. The decision maker is able to watch the progress of the model through time and has the ability to interact with it.

The current levels of interaction are shown in Fig. 2.

The example shows the simulation has reached TIME = 57. It also shows two queues; the first described as JOBQUEUE has the entities JOB.1, JOB.2 and JOB.3 queueing at it, while the second queue, MACH.1, has the entities JOB.4 and JOB.5 queueing at it. The different levels of interaction are:

(i) *Basic operations.* This level of interaction is called basic because it contains operations which make one-to-one changes in the model. This includes the ability to delete, duplicate or move any entity that is displayed on the screen. The changes are one-to-one since any change on the screen gives a similar change in the model.

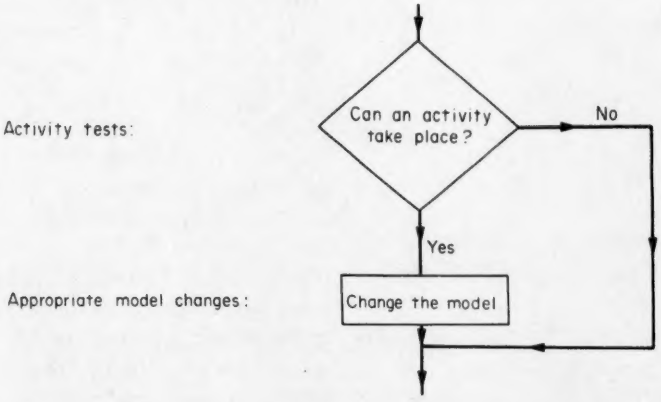


FIG. 1.

Vol. 6
1978

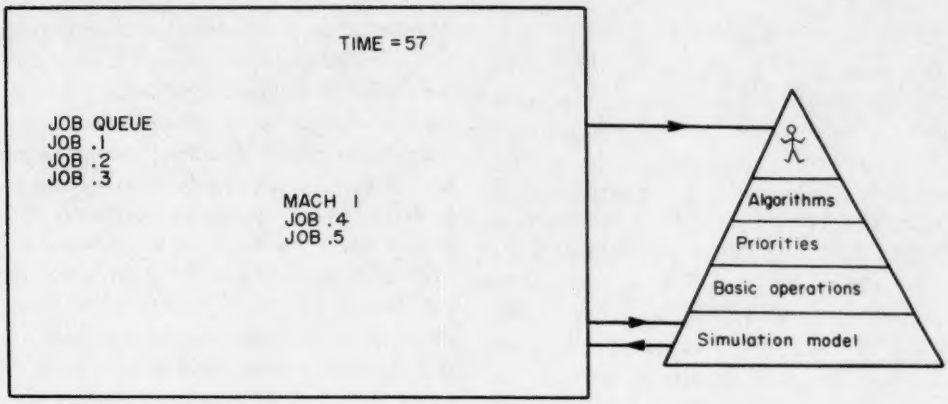


FIG. 2.

Since it is possible to change any entity or any attribute of any entity, then the use of these basic operations, together with the activity scan approach, can completely change the structure of a model at run time.

(ii) *Priorities*. The basic operations form a powerful set of instructions by which the decision maker can alter the progress or structure of a model. This approach is time consuming if used frequently by the decision maker.

A common basic operation is that of giving priorities to operations, jobs or entities. A macro version of this was designed which gives dynamic priorities to entities for the remainder of a simulation run.

(iii) *Algorithms*. The highest form of current interaction is contained in the different driving algorithms. The analyst would design separate algorithmic modules for his particular model which the decision maker would then be able to link together to drive a specific simulation.

A SAMPLE PROBLEM ANALYSED BY VISUAL INTERACTIVE SIMULATION

Visual Interactive Simulation has been used to explore several types of problems. One example problem is described in this paper. The aim in introducing this sample problem is to describe the technique of visual interactive simulation. The problem which is described below relates to the operations of a simplified coal yard.

The problem consists of two types of vehicle, known as NCB and MERCH, which arrive, randomly, at a weighbridge. After weighing in the NCB vehicles remain in the yard for a fixed amount of time doing specific tasks before filling up with coal, and weighing out. The MERCH vehicles, which also arrive randomly, need first to be weighed in. After this operation they proceed to a loader, where they are loaded with coal and from here leave the yard after having weighed out. Coal trains, known as TRNs bring coal into the yard, are unloaded using the same loader required for the MERCH vehicles, before they too leave the yard. The situation requires investigation since:

(a) alternative priorities at the weighbridge and loader can be in operation which will affect the efficiency of the yard; and

(b) given the data on arrival patterns of the vehicles and trains unstable situations occur.

Whilst the data and logic appear as simplifications of reality, the exercise is nontheless illustrative of a general type of routing/scheduling problem, where entities within the model are queueing for scarce resources and services. In batch simulations, conflict over the order in which entities are allowed to pass through the system must be resolved before run-time, by drawing up a set of priority rules. By using a visual interactive simulation approach it was hoped that this conflict could at best be minimised, at worst understood. Appendix 2 shows the logic for the problem together with some sample output. A similar problem is described by Jones and Ryder [6].

(1) *Conflict area—weighbridge*

Both types of lorry queue at the weighbridge, initially in order to weigh in and then subsequently in order to weigh out. Several priority options are open to an analyst in a batch mode simulation. The analyst could specify an exact schedule (e.g. NCB in, NCB out, MERCH in, MERCH out) or could use a LIFO or FIFO schedule. However, in a real-life situation, it may often be better to operate more flexibly, and change scheduling policies as the current situation in the depot appears to warrant. For example, if a large number of lorries were already queueing within the depot, it is likely that the weighbridge operator would give priority to lorries wanting to weigh out. Conversely, if equipment were standing idle within the depot, he would probably give precedence to those coming in. In an interactive model it is possible to change priorities dynamically.

(2) *Conflict area—loader*

The services of the loader are sought by two types of vehicle. Trains use the loader to off load coal while merchant lorries use the loader in order to load up. A batch simulation would require a pre-determined schedule of priorities for the use of the loader.

(3) *Conflict area—coal stocks*

Both types of lorries are competing for coal stocks when they arrive within the depot. Naturally when stocks are high, the importance of their conflict is negligible. However, as stocks dwindle, the decision "Who can take up stocks?" is not insignificant. A batch approach forces the analyst to decide how to distribute coal between NCB and MERCH lorries before run-time. In the light of experience, a manager might be able to point to unrealistic aspects of the analyst's schedule, and suggest changes which could be implemented on-line, using an interactive system.

What are we able to learn from watching this simple demonstration program progress through time under a FIFO schedule?

The first point to note is that having a visual representation of a simulation model aids the verification of that model. It is possible to step the model forward either by one simulation time unit or to the next event. This method of time advance, linked with the ability to interrogate attributes and entities enable the logic of the simulation model to be checked. Further, by setting attributes associated with interarrival or activity duration times arbitarily high, it is possible to investigate how the model behaves under extreme conditions. These actions aid the verification of the model.

It soon became evident from watching the simulation progress through time that the major problem was that of low coal stocks. If stocks diminish to zero, large queues of lorries build up in the depot, all waiting for coal. Not until a train arrives and off-loads can any of this backlog be cleared away. However, this event in turn causes a second-stage problem. As lorries take up these coal stocks then a heavy demand is now transferred to the weighbridge. Under a FIFO schedule, this meant that, whilst the weighbridge was handling lorries trying to leave the system, lorries wishing to enter were jammed outside, causing underutilisation of machinery in the depot.

It was clear that the visual aspect of the system in operation enables us, as analysts, to locate the practical problem areas (i.e. critical bottlenecks), and to reach conclusions which would not have been immediately self-evident from running batch experiments. If several full trains arrived early in the simulation (as in the GOOD example in Appendix 2), no problems arise—stocks were kept high and were easily capable of matching demand, which resulted in a stable equilibrium. However, if little coal was seen to arrive in the first part of the experiment (see the BAD example), problems were enormous. The two-stage, 'pendulum' bottleneck effect described above threw the depot into chaos.

With a method of visual simulation, we are able to identify reasons why the system is not efficient, and does not operate as we would like (in this case, we can see that the high variability of both train arrivals and coal capacities are mostly to blame). From this, we can obtain heuristically-developed answers to questions of how to improve operations.

THE CHOICE OF TERMINAL

The initial work of investigating visual simulation methods at Warwick used a computer graphics terminal [9]. The advantage of this

type of terminal over the more common alpha-numeric visual display unit, is that is offers full graphics facilities, a far better resolution and a refresh display method of drawing pictures on the screen. Computer graphics terminals typically have resolutions of the order of 1000 lines by 1000 columns within a screen size of 10″ by 10″. The resolution of the screen, linked with the fact that the picture is re-drawn 25 times every second, makes it a relatively easy task to 'move' items on the screen simply by re-transmitting its X and Y co-ordinates.

The major disadvantage of using a computer graphics terminal has been that, until relatively recently, this type of device has been non-portable. All managers/decision makers had to visit Warwick in order to use the approach.

The use of an alpha–numeric visual display unit has made the approach portable. The main disadvantage, however, is the lack of resolution available on this type of terminal. A typical visual display unit consists of 20 lines by 80 columns and in order to make an item 'move' on the screen it must first be deleted from its current position. This action is followed by a second message to re-display the item at its new co-ordinates. Thus the number and length of messages transmitted from the computer to the terminal in order to make an item move is inefficient.

Some current work at Warwick is investigating the use of intelligent, colour terminals for visual simulation. This type of device is portable, allows graphical output and is programmable. When using this type of terminal only condensed display messages need be sent between the main simulation program and the terminal.

CONCLUSIONS

This paper has introduced the ideas of visual interactive simulation that have been developed at Warwick.

It is interesting to compare this approach with the work described by Palme [8]. He gives examples of 'moving picture' simulation models developed using SIMULA [2]. His approach is to build a simulation model in the normal batch mode. After the batch model has been developed the visual and interaction facilities are then included. This article, however, suggests that visual and interactive facilities are also important at the verification stage of a simulation study.

The technique, used at Warwick, is to build a visual interactive model where the visual and interactive facilities are an integral part of the model development. This approach has been found to assist both the analyst in developing the model and the manager/decision maker who will use it to explore his problem situation.

REFERENCES

1. CLEMENTSON AJ (1973) *Extended Control and Simulation Language: Users Manual.* University of Birmingham, The Institute for Engineering Production UK.
2. DAHL O & NYGSAND K (1966) An ALGOL-based simulation language *Comm. A.C.M.* **9**(9).
3. FISHMAN GS (1973) *Concepts and Methods in Discrete Event Digital Simulation.* Wiley, London.
4. GORDON G (1969) *System Simulation.* Prentice-Hall, Englewood Cliffs.
5. HILLS PR (1967) *A Computer Language in Algol in Digital Simulation in Operational Research* (Ed. HOLLINGDALE SH), English Universities Press, London.
6. JONES L & RYDER P (1975) *Simulation Modelling.* T341 Module 6, The Open University Press, Milton Keynes, UK.
7. MATHEWSON SC (1974) Simulation program generators. *Simulation* **23**(6), 181–189.
8. PALME J (1977) Moving pictures show simulation to user. *Simulation* **29**(6), 204–210.
9. *The Graphical Display System 4100* (1969) International Computers Limited, Putney, London, UK.

ADDRESS FOR CORRESPONDENCE: *Dr RD Hurrion, University of Warwick, School of Industrial & Business Studies, Coventry, Warwickshire CV4 7AL, UK.*

APPENDIX 1

VISION—A visual interactive simulation system

The visual interactive simulation system is primarily an actively based simulation system. The language is similar to SIMON [5]. It consists of a set of ALGOL procedures which are called by an ALGOL simulation program.

These procedures fall into four categories:

(1) *Initialisation procedures*, which start the model, and which define the size and nature of the simulation under consideration.

(2) *Visual display procedures*, which help to display the model on the VDU screen.

(3) *Simulation procedures*, which affect both the ongoing state of the model and the way in which it is depicted on the screen.

(4) *Interaction procedures*, which help the analyst to alter both the simulation model and visual display at 'run-time', i.e. to affect the course of the simulation as he watches it take place.

APPENDIX 2

Monitoring of the coal depot experiments

The description, given earlier, of using the visual system on the sample problem stated that the instability of the system could be seen to be caused by the variability of coal arrivals. To illustrate this point, the model was run several times, and 'snapshots' of how it changed through time were taken. The following pictures are copies of what appeared on the visual display unit at the times shown in the top right-hand corner of the screen. Stocks of coal at that point in time are in the bottom left-hand corner: one 'C' represents a lorry-load of coal. Vehicles move around the screen as shown in Fig. 3.

Appendix 2.1

A good run

During this experiment, significant quantities of coal arrived early at the depot. Figure 4 shows that, at time 92, fourteen lorryloads of coal were in stock. Further snapshots at times 291 and 570 (Figs. 5 and 6) indicate that

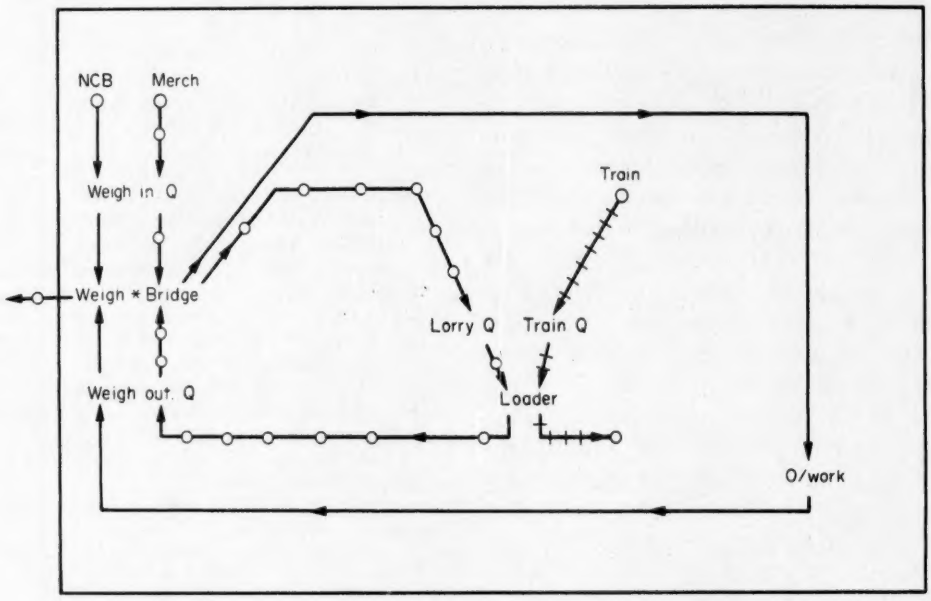


FIG. 3.

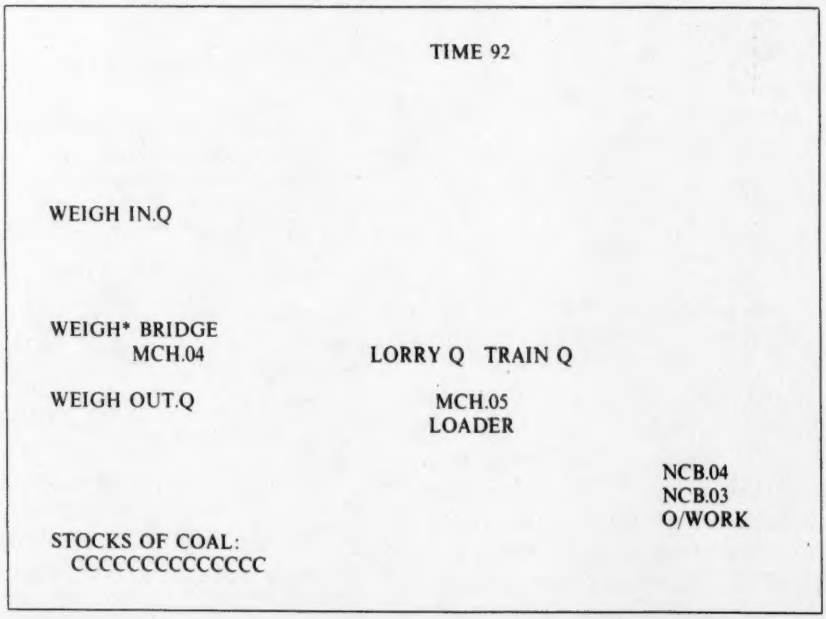


FIG. 4.

Vol. 6
1978

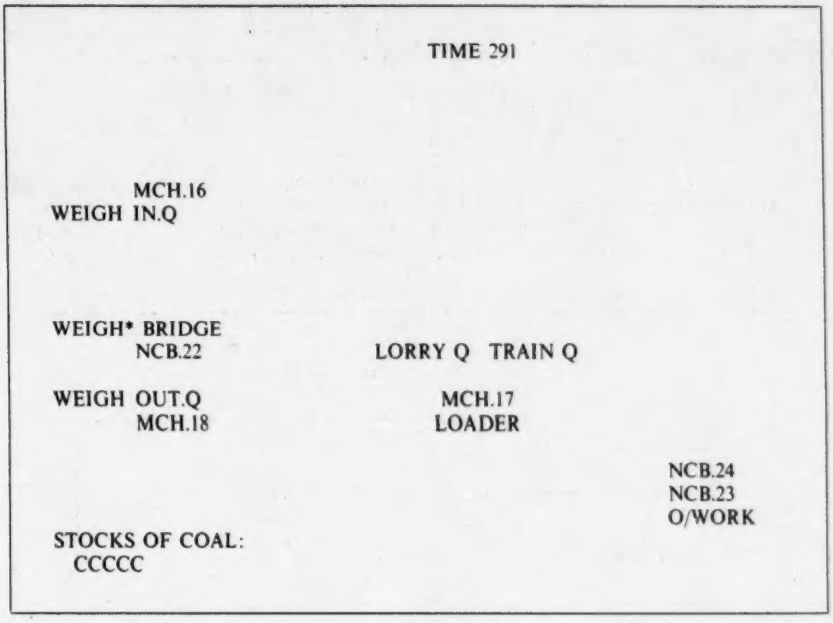


FIG. 5.

the effect of this early, large stock of coal leads to stable, smooth situation within the depot.

Appendix 2.2

A bad run

This experiment, on the other hand, was typical of the stochastic nature of simulation runs, when little or no coal arrived early on. Notice that at time 101 (Fig. 7) stocks of coal are zero, and already queues of MCH and NCB lorries are building up. Because of the NCB lorries' priority over new stocks of coal, the MERCH queue soon causes chaos within the depot (see time 535, Fig. 8).

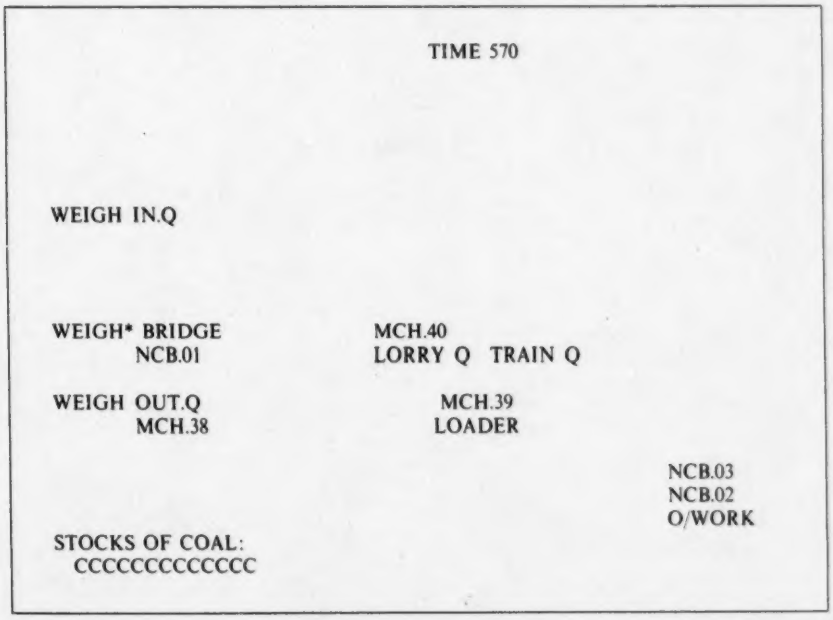


FIG. 6.

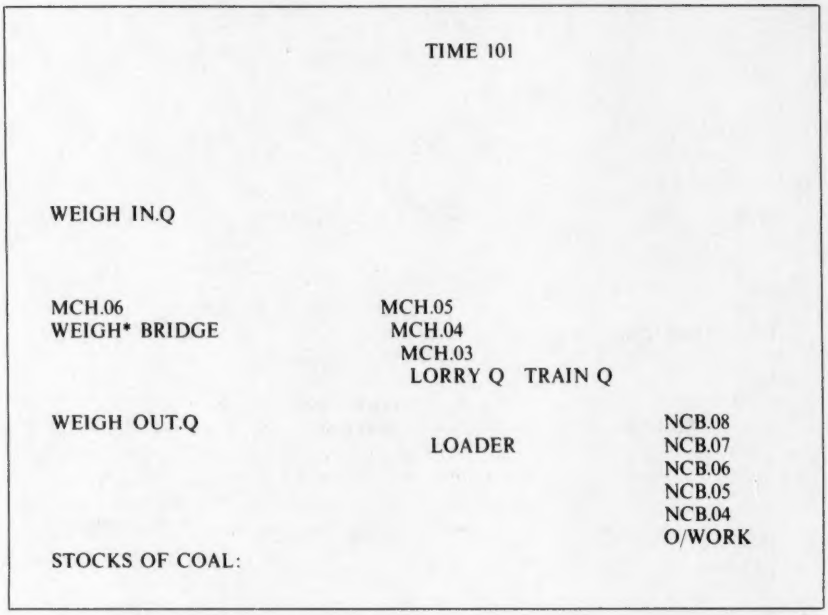


FIG. 7.

Vol. 6
1978

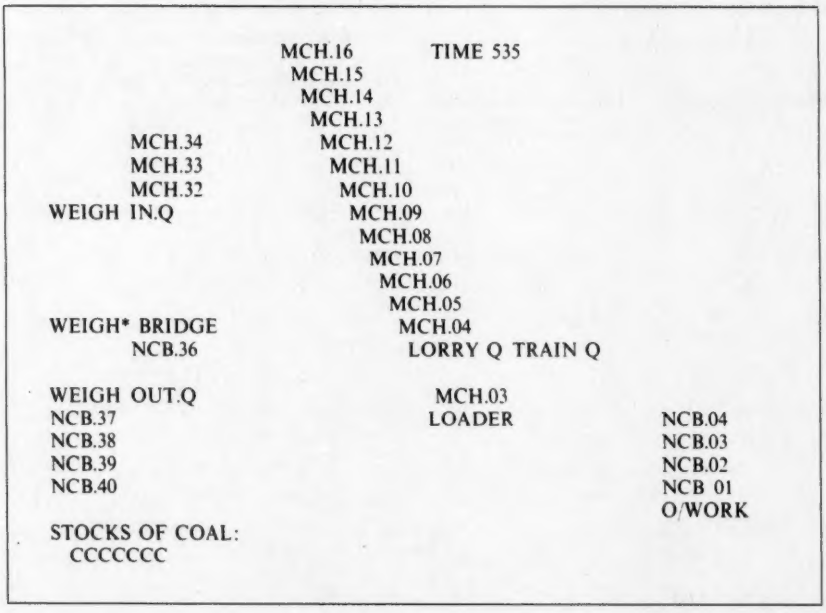


FIG. 8.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 6, No. 5, pp. 427–431
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0427$02.00/0

Planning with Multiple Objectives

PAUL WIEDEMANN

Georgetown University, Washington, D.C.

(*Received March* 1978)

Planning any task that is characterised by the presence of multiple objectives presents the decision maker with a set of special considerations which are not present in the case of decision making with one unique objective. The paper outlines the stages in the formulation of a dynamic multiple-objective objective function and introduces the criterion complex as a decision making construct particularly suited to formulating and solving problems with multiple objectives. The paper then considers some of the mathematical properties of the alternative formulations of criterion complex.

I. INTRODUCTION

1.

THE TRADITIONAL form of representation of the development process of a system—be it a national economy, a private corporation, or a hospital—was as a uni-dimensional system with a single objective—maximizing the growth of GNP, maximizing profit, etc. With time the other objectives became increasingly important—income distribution as well as the growth rate, sales volume as well as profit—and the development of mathematical programming introduced much more sophisticated ways of handling such multidimensional objectives (expressed in the form of both objectives and constraints).

The simultaneous development of the computer and of computer programmes for LP made the solution of such problems easier. And the refinement of LP programmes and the increase in the size and speed of computers has made it possible to solve highly sophisticated and quite large problems quite quickly and inexpensively.

2.

But there has not been a parallel development in the theory and application of multiple-objective objective function. And this is despite the fact that it has become clearly recognised in more and more spheres that this is the nature of the problem which must be solved —be it in finance, industry, land use, education, the health service, subsidies to the arts or whatever. And as this has been accepted, decision makers have come to see, as Peter Wiles has said, that "the most difficult task faced by decision makers is not problem solving but problem formulating" [5, p. 264]; and, moreover, that the task of problem formulating becomes increasingly more difficult as the system becomes more complex.

3.

At the general level, at the level of the *content* of the criterion, much has be written about the nature of the objectives of the decision making process in various spheres and at various levels. But appreciably less work has been done at the more concrete level, at the level of the specific *economic and mathematical form* of the criterion. Yet it is only at the level of the economic form that the criterion appears sufficiently refined to be measured, since a given content allows numerous formulations.

It has been said that in the modelling of economic systems "econometricians have typically left the definition of the appropriate welfare or objective function to the welfare theorists, who have generally ignored the problems of definition and measurement" [1, p. 9]. It is on the economic form of the multiple-objective objective function that we will focus here; but if the criterion thus formulated is then to actually be used in modelling, measurement

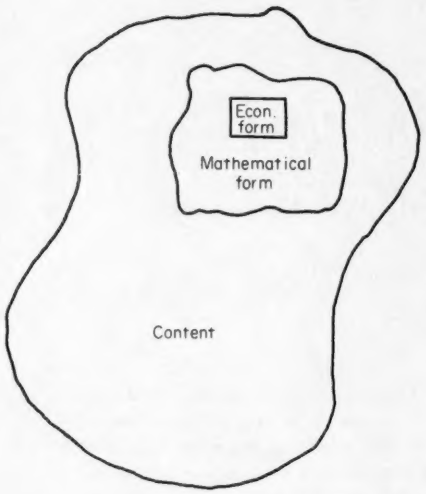


FIG. 1. *Content, economic form and mathematical form of the criterion.*

must be carried yet further in the specification of the mathematical form of the criterion. (Figure 1 shows the relationship of the content, economic form, and mathematical form.)

II. FORMULATING A MULTIPLE-OBJECTIVE OBJECTIVE FUNCTION

1.

As a starting point of the discussion, we assume that multiple objectives do in fact exist and that this is the best way to formulate the problem in question. This, of course, is a very large assumption and indeed one that should not be firmly stated at the outset of the evaluation process, but which must be re-examined as the results of the evaluation procedure become known.

More specifically, this initial assumption is that (some or all of) the multiple objectives cannot be collapsed into one another and the problem reformulated in a simpler formulation. And that the problem cannot be reformulated with (some or all of) the multiple objectives reformulated as part of the constraints on a simpler objective function. Finally, it is assumed that no important objectives have been omitted.

2.

Given that we have a multiple objective problem, one must then consider whether the objectives can be both readily measured and summed in some agreed manner. (As would be the case, for example, where the problem was concerned with measuring several dimensions of cost; but not necessarily where one was considering several different aspects of social welfare.) In the case where objectives can be measured and summed, and where no important objectives have been omitted, we can proceed directly to the problem of inter-temporal weighting discussed in 6 below.

Increasingly, however, and especially as one attempts to address problems outside the traditional private economic cost and revenue domain one can no longer assume that the results of the process of quantification for formulating the multiple objective function are correct in every respect. Specifically, debate proceeds on the questions of, firstly, whether the objectives can be measured; and secondly, whether, once measured, the multiple objectives can be expressed in a comparable unit of measurement in the objective function.

3.

If no reasonable quantification exercise appears possible, resort must be made to an ordinal method of evaluation. Except in the one rare case where one scheme was unequivocally ranked first on all criteria, the information provided by such a ranking of disparate items of advantage and disadvantage "is no more advantageous than no information at all" [2, p. 95].

An ordinal method of evaluation identifies the fact that a multiple-objective objective function exists, but it does not fully identify the conflicts within the function, let alone resolve them. It therefore appears desirable to attempt to design some process of quantification—however elementary and crude—in order to further the process of evaluation of alternatives and thereby to strengthen the role of quantitative planning methods as an aid in the decision making process.

4.

Accepting the need for cardinal evaluation of alternative structures of objectives in a multiple-objective decision making framework, attention must then be focused on designing a method of quantification of objectives which would be accorded a minimum level of accept-

ance by the decision makers. But any such proposed evaluation scheme must be critically examined, great attention being paid to marginal changes in the value of the function as a result of changes in the measures of variables used and to marginal changes generated by deleting old variables or introducing new ones. Or by removing or introducing constraints on the function.

If these changes produced relatively major changes in the value of the objective function, this would point to variables or constraints which needed to be further examined with regard to their inclusion, exclusion, or proper formulation.

5.

Having quantified the elements of the multiple-objective function, the next step is to design a system of inter-elemental weights. This again is more difficult for problems outside the traditional private economic cost and revenue domain where the answer to the question of 'whose preferences decide the weights' is relatively straightforward (though less so in the framework of the modern corporation than in the small firm).

In the social and political sphere the question of 'whose preferences' makes inter-elemental weighting a much less objective—and much more sensitive—problem. This means that if minor changes in these weights produced relatively major changes in the value of the objective function, the weighting scheme would clearly need to be further examined. In carrying out the process of evaluation of alternative formulations of the multiple-objective function, the planner would be under an obligation to carefully specify to the decision makers the basis of the inter-elemental weightings underlying the alternative formulations of the function.

6.

With an agreed set of elements for the objective function and an agreed set of inter-elemental weights, the final problem area is that of inter-temporal weighting of the objective function. Unless the problem faced is of a one-off, non-recurring nature, it is necessary to relate solutions in one time period to solutions in each of the other time periods of the planning period. This involves an explicit consideration of the inter-temporal interdependency of solutions in successive years.

Each coefficient of the objective function will therefore be the product of two separate weighting factors: one expressing the inter-elemental weighting of the elements one against the other, and a second expressing the inter-temporal weights for each year of the planning period. In the final form of the criterion function for any multi-period planning process, the two weights will lose their separate identity and will both be expressed by the coefficient to the variable in the criterion function.

This six stage process of the formulation of a multiple-objective objective function is represented diagrammatically in Fig. 2.

III. FURTHER DEVELOPMENT OF THE MULTIPLE-OBJECTIVE OBJECTIVE FUNCTION

1.

In moving from the more general content of the criterion to the more specific economic (and mathematical) form, it becomes progressively clearer that the so-called 'objective function' of the decision maker is not something objective at all, but something essentially subjective. The objective function expresses the preferences of the decision makers among a set of objectives—and both the preferences and the objectives themselves are changing. Moreover, the decision makers themselves are generally "an ensemble, a coalition as it were, of groups which are more or less in conflict with each other" [3, pp. 33–4, 4].

2.

It is our view, therefore, that a more preferable term to the objective function is the criterion function. But more importantly, it is basic to the programming approach to decision making that importance is attached not only to the criterion function, but to the constraints on this function as well. To stress the importance of the constraints often relatively or even absolutely neglected in the theoretical literature—as well as of the criterion function, the term *criterion complex* should be introduced to encompass the criterion as well as its constraints.

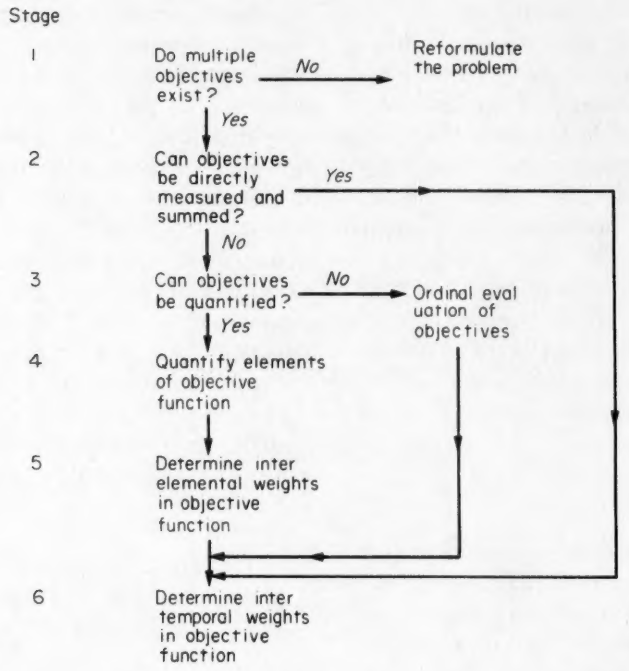


FIG. 2. *Stages in the formulation of a dynamic multiple-objective objective function.*

3.

The process of formulating the criterion complex is a sequential arrangement. For many economic planners it has been assumed to be a two-stage process in which one first determined the limitations and then one formed the criterion of optimality. For others, particularly those believing in the objectivity of a certain criterion function, the stages are reversed.

In fact the specific economic and mathematical form of the function depends on the elements in the constraints. And, as exogeneous political, economic, social, institutional, definitional, or technological changes change the constraints, so must the form of the function be re-examined. The two are shaped together in an iterative process and they are both part of one whole—the criterion complex.

4.

The process of the formulation of the multiple-objective objective function discussed in section II and diagrammatically expressed in Fig. 2 can now be seen to be part of the larger process of formulating the criterion complex, the stages of which are represented in Fig. 3. In this context, the formulation of the multiple-objective function of Fig. 2 would be part of stage 1 of the formulation of the criterion complex of Fig. 3.

5.

The importance of the idea of the criterion complex is brought home by the distinction between a technical system and a social system: a technical system has a goal that is quite general, that has no relationship to the system and to the elements within which it functions. When we consider a social system (as we do of course in modelling not only the educational or health sectors but also a modern corporation) there are certain bonds between the object, the existence of the system, and the existence of its elements. A social system has as part of its object the attempt to maintain its existence, and not only its existence as a whole, but also the existence of the elements of which it is composed.

Whereas in the case of the technical apparatus any structure is suitable if its behaviour makes possible the advancement of the desired properties of the system, the demand made on the structure of a social system is to achieve

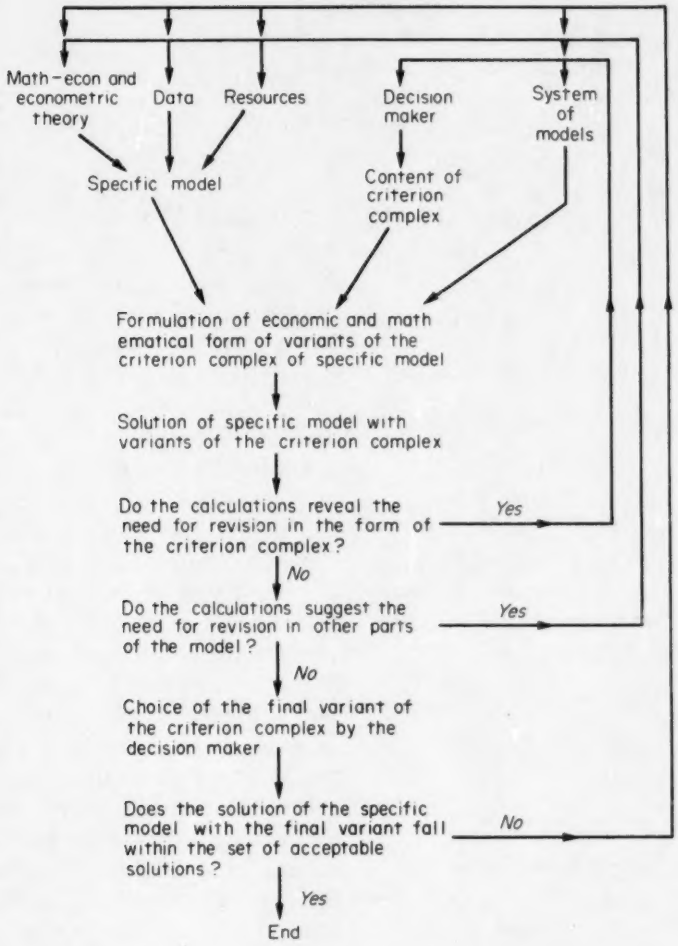

Math-econ and econometric theory
Data
Resources
Decision maker
System of models
Specific model
Content of criterion complex
Formulation of economic and mathematical form of variants of the criterion complex of specific model
Solution of specific model with variants of the criterion complex
Do the calculations reveal the need for revision in the form of the criterion complex?
Yes
No
Do the calculations suggest the need for revision in other parts of the model?
Yes
No
Choice of the final variant of the criterion complex by the decision maker
Does the solution of the specific model with the final variant fall within the set of acceptable solutions?
No
Yes
End

FIG. 3. *Formation of the criterion complex.*

certain economic results, but at the same time observing the wishes of members of society with regard to certain aspects of their environment.

6.

The relationship between the criterion function and the constraints in the criterion complex perhaps becomes clearer in a mathematical example. The objective of the management process can be expressed (in linear form for simplicity)

(1) $Z = \tilde{p}\,\tilde{x} \rightarrow \max.$

(2) $A\,\tilde{x} \leqslant \tilde{b}$

(3) $B\,\tilde{x} \in D$

Here $\tilde{x}$, $\tilde{b}$ and $\tilde{p}$ are vectors of outputs, inputs, and value weights (prices), respectively, and A and B are input–output matrices. The condition $B\,\tilde{x} \in D$ reflects all the constraints of a social nature on the management process (not included in (2)).

In a technical system the criterion complex would have been (1), whereas in a social system the complex is expanded to also include at least part if not all of (3).

Solutions to the problem ((1) + (2)) may well not satisfy the broader problem ((1) + (2) + (3)). Indeed, a solution of the former may suggest that some of the elements of (3) should become additional elements in the multiple-objective objective function (1) or vice versa.

7.

Expressing (1) as:

(1)a $Z = p_1x_1 + p_2x_2 + \ldots + p_nx_n \rightarrow \max.$

the $a_1 s$ are the *inter-elemental criterion weights* of 11.5 (above). Reduced to only one element ($Z = p_1 x_1 \rightarrow \max.$), the objective function could be called a *uni-dimensional criterion* function, and (1)a a *multi-dimensional criterion function* (ie. a multiple-objective objective function).

By adding a weighting factor $\phi(t)$, and moving from the linear formulation of 111.6 and 111.7 to a more general one, the criterion function becomes:

$$Z = \phi(t)\, s\,(\tilde{x}^t) \rightarrow \max.$$

This is the *inter-temporal criterion function* of 11.6 (above).

8.

The inter-temporal aspect of the criterion function could be further developed by considering a *terminal criterion function*, this being one where the optimisation process is focused on the horizon year T. Thus:

$$Z = s(x^T) \rightarrow \max.$$

This could be contrasted with a *cumulative criterion function*:

$$Z = \Sigma_{t+1}^{T}\, \phi(t)\, s\,(\tilde{x}^t) \rightarrow \max.$$

In this formulation it would no longer be possible to neglect the values of the variables over the interval $(t + 1)$ to $(t + T - 1)$. The extent of the difference in the solution for Z will depend upon the nature of the constraints imposed on the solution, on the nature of the weighting function, and on the mathematical form of the function.

9.

Attempting to determine the proper nature of the cumulative function one must consider whether it should be *continuous* or *discrete*. That is to say, should it be the integral over the period, or can it be correctly represented by the sum of a series of values for separate intervals of the time period? Here practicality compels us to opt for a discrete formulation.

Finally, we have the choice between an independent and a memorising criterion function. An *independent criterion function* is one of the traditional form. It can be contrasted with a *memorising criterion function* where the maximum depends not only on the present values of the state variable, but also on its value in a preceding period:

Thus

$$Z = s\,(\tilde{x}^T;\ \tilde{x}^{T-1})$$

or

$$Z = \Sigma^T\, \phi(t)\, s\,(\tilde{x}^T;\ \tilde{x}^{T-1})$$

Such a function represents a step away from focusing on one time period, and a move toward the incorporation of a wider view of the continuous nature of the planning and management into the programming approach to the formulation of the planning and management process.

The argument of this paper is that any planning and management activities which it is suggested are characterised by multiple objectives should be analysed following the steps of section II as part of the process of determining an initial set of operable planning objectives. In dynamising the planning process it is furthermore necessary to make explicit recognition of the fact that objectives are not in fact objective, and to formulate the planning objective at any stage of the development process as part of the criterion complex, including both the criterion function and the preference constraints, of the organisation. With the objectives set in a dynamic framework it is then necessary to consider the precise nature of the relationship of objectives over time—are they cumulative, independent, etc.—in order to complete the specification of the dynamic economic form of the multiple objective objective function.

REFERENCES

1. Frisch R (1963) An implementation system for optimal national economic planning. Mimeographed. Sosialøkonomiske Institut, Oslo, Norway.
2. Kettle P & Whitbread M (1973) An ordinal method of evaluation: a comment. *Urban Stud.* **X**, 95–99.
3. Kornai J (1971) Economic systems theory and general equilibrium theory. *Acta Oeconomica* **VI**/4, 297–317.
4. Wiedemann P (1973) A critique of the planners' preference function. *Osteuropa Wirtschaft* **XVIII**/3–4, 244–246.
5. Wiles P (1967) Some fundamental questions on national planning in (Gross B) *Action Under Planning: The Guidance of Economic Development*, pp. 264–280. McGraw-Hill, New York.

ADDRESS FOR CORRESPONDENCE: *Dr Paul Wiedemann, Department of Economics, Georgetown University, Washington, D.C.*, 20057, USA.

OMEGA, The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 433–442
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0433$02.00/0

Work System Design and the Quality of Working Life

DAVID W BIRCHALL[1]

Henley, The Administrative Staff College, UK

(*Received February 1978; in revised form April 1978*)

Those responsible for designing work systems have considerable influence over the quality of working life of those employed both in operating and managing the system. Whilst research evidence is available regarding the impact of changes to jobs and work organisation, this research has tended to be based on changes made for remedial reasons where existing arrangements have been seen as the source of problems such as high labour turnover, low productivity, etc. The changes reported have often been limited in extent since assumptions regarding aspects such as the operations technology remain unchallenged. Those responsible for the initial design of work systems, however, have an opportunity to consider means for developing a "fit" between the needs of employees and the objectives of the organisation throughout the design process. Here we shall consider those aspects of work systems which relate particularly to the motivation of employees. Two situations are described, and contrasted, where those responsible for the design of the work systems place emphasis upon integrating the needs of individual employees into the initial design.

THE WORK system designer is required to develop a system for converting inputs into outputs in order to achieve certain agreed objectives. Technological options are often limited; considerations of technological feasibility and economic viability tend to dominate the decision regarding the basic operating technology. Whilst those responsible for design are required to make early decisions regarding the basic operating technology in order to obtain financial commitment, the detailed design of the managing and operating system is usually deferred. Even though technological decisions put constraints upon these subsequent choices, in most cases organisational options will still exist.

The organisation must create those conditions necessary for job-holders to be motivated to achieve the tasks required to satisfy organisational objectives. Therefore, at the stage of detailed design the work system designer is attempting to develop a work organisation which meets organisational objectives, both economic and technological, as well as the needs of individual employees both psychological and social. Longer-term, if employees are to be retained within the system, the organisation must be capable of satisfying the changing needs of job-holders. Whilst the cost of replacing dissatisfied employees who leave the system can be estimated (recruitment, induction, training) prediction of labour tenure is less certain, as is the cost associated with demotivated employees engaged within the system. Organisations capable of adapting to the changing needs of both the organisation and the employees may be seen as desirable. For this to be achieved the work system designer has to be able to forecast these changing needs in addition to establishing current needs and expectations of potential employees.

The psychological requirements of individuals relating to job content have been suggested by Emery and Thorsrud [4] to include:

1. The need for the content of the job to be

[1] The author is supported by a research grant from the Social Science Research Council.

reasonably demanding in terms other than just physical effort and yet providing some variety.

2. The need to be able to learn on-the-job and develop personally through the job.

3. The need for some area of decision-making that the individual can call his own.

4. The need for support and recognition at the workplace.

5. The need to feel that one's job is acceptable to others outside the workplace.

6. The need to feel that the job leads to some sort of desirable future.

It is clear that individuals differ in their psychological need strength and that some workers are satisfied with jobs appearing to others as highly repetitive, low in skill demand and responsibility. The designer wishing to integrate individuals with such differences may aim to design a work organisation providing flexibility in jobs so as to enable each individual the opportunity to satisfy his own needs.

In designing a work system to meet the needs of individuals it is necessary to consider the tasks required and the aggregation of tasks to form work as allocated to separate job holders. It is also necessary to allocate decision-making between different members of the organisation. Decisions can be classified into both *control decisions* and *regulatory* decisions. Control decisions specify the outcome objectives or targets and deal with both the boundaries between the work system and its external environment (*boundary decisions*) and the method for transforming inputs to outputs (*intra-system decisions*). Boundary decisions cover such aspects as resource planning, specification of quality and quantity, the scheduling of inputs and outputs, whereas intra-system decisions specify procedures and methods for achieving the outcome objectives. Regulatory decisions are those which relate to the control and regulation of the process. These involve monitoring the conversion process and decisions taken to maintain the process within previously specified performance limits.

Traditionally, the work system designer has been required to specify precisely the methods to be employed to achieve system objectives.

Such detailed specification of methods of working creates tightly defined jobs for individuals, hence limits the flexibility available for these individuals to create their own role within the work group to meet their own personal needs. Though this approach is intended to lead to the "best method" it may prove dysfunctional. Employees may be faced with the situation where the "rules" tend to hinder task performance and concerted efforts to "work to rule" may lead to non-achievement of system objectives. Satisfactory performance may only occur when experienced operatives develop and operate their own system.

Cherns [3] advocates as a principle for the design of organisations the minimal critical specification of the tasks and the minimal critical allocation of tasks to jobs or of jobs to people. Specification of objectives remains essential but the means for obtaining them in many instances can be decided by the task performer. This approach may result in a greater degree of flexibility for individual job holders within the work system and allow for direct employee involvement in both control and regulatory decisions.

Attributes of jobs which contribute to the motivation of employees [1] have been identified as follows:

1. the level of variety

2. the extent of repetitiveness

3. the degree of attention with accompanying mental absorption

4. the level of responsibility for decisions and the discretion present

5. employee control over his own job

6. the presence of goals and achievement feedback

7. perceived contribution to a socially useful product

8. opportunities for developing friendships

9. where dependent upon others for task

achievement some influence over the way the work is carried out

10. perceived skill utilisation

It is now necessary for the work system designer to consider to what extent the activities required to achieve work system objectives can be structured so as to provide these attributes in jobs, the need strength of potential employees for each attribute and the extent of specification appropriate. Since much of our focus has been upon the design of work systems, many of which require a team rather than individual effort, it is also necessary to consider principles for the design of work group activity. Guidelines incorporating the principle of minimal critical specification [6] include:

1. Primary work groups should include between 4 and 20 individuals.
2. The primary work group should have a designated leader who is accountable for the group's performance.
3. The group should be responsible for activities which make-up a complete task.
4. Wherever possible the group members, through their supervisor, should be responsible for planning their own work.
5. Members should have the opportunity to evaluate their performance compared with the standards.
6. All the skills necessary for satisfactory task achievement must be present within the group.
7. Reward systems should be designed to support group behaviour.

The degree to which job attributes are present which contribute to employee motivation appears to be closely related to the nature of the tasks to be performed, the extent to which control and regulatory decisions are required on an on-going basis and the degree to which those actively engaged in task performance directly influence these decisions. Both the nature of the tasks and the decision-making requirement are related to the technology employed for the transformation process, e.g. in an automated process the control of the process may be carried out by machine. Additionally, the nature of the market for the product or service will influence the degree of adaptability required which in turn will influence the nature and extent of decision-making relating to the work system, e.g. a large and stable product market enables long product runs and hence less frequent but more significant boundary decisions are required. Additionally, if the unit for analysis is individual work elements, the scope for decision-making may be considerably restricted when compared with that essential for the operation of the work system. Consideration of the total work system and the design of the work group activity appears to be the more appropriate level of analysis.

In order to design the role for the work group and its relationship with management the first step is to identify those decisions which are necessary to operate the work system. The typical decisions are listed in Table 1 under the categories identified earlier, i.e. boundary control, intra-system control and regulatory decisions. The regulatory decisions are concerned with the control of system variances, i.e. unprogrammed occurrences which affect conversion efficiency. Cherns [3] advocates that these variances be dealt with as near to their point of origin as possible in order that corrective action can be taken with minimum delay. For this to be possible those responsible for regulation need information about the process, knowledge of the nature of variances and an understanding of the corrective interventions and their impact. Monitoring may be constantly necessary during the conversion process and regulatory decisions and action frequent. Control decisions are typically made less frequently, some being required once only. Where the control of variances is to be undertaken as near to the point of origin as possible and by the employee in closest proximity, involvement of that employee in intra-system decision-making is logical where either local knowledge is essential or the individual has the capability and interest in participating.

TABLE 1. DECISIONS REQUIRED IN OPERATING SYSTEMS

CONTROL DECISIONS

A. *Boundary Control*
1. Goals—qualitative (what to produce and when)
—quantitative (how much to produce, the standard of quality and the cost)

2. Resource availability and acquisition (internal & external to work system)
human
materials
equipment
space

3. Management of boundary interface
—information flows (inputs, outputs, resource utilisation)
—decision-making authority and responsibility
—specialist role

4. System performance
—criteria for assessment (human, material, equipment, space)

B. *Intra-System Decisions*
1. Working methods

2. Time utilisation and allocation of tasks

3. Work-group leadership

4. Relative status of work-group members

REGULATORY DECISIONS

1. Materials
—quality control of incoming materials & components
—maintenance of supply

2. Equipment
—set up and adjustment
—request and authorisation of maintenance work
—authority to cease operation
—allocation of repair work/maintenance

3. Labour
—select and appoint
—disciplinary action
—discharge
—skill assessment and development

4. Product
—output quality
—output rate
—reject levels
—rectification/repair defects

5. Work area
—cleanliness of work place/area
—safety at work place/in area

6. Conversation process
—scheduling work-flow
—controlling variances

WORK SYSTEM DESIGN—CASE EXAMPLES

In this section two cases are presented which illustrate different approaches to the problem of integrating the needs of individual employees and organisational objectives in the design of work systems. The stage in the design process at which significant attention was paid to the design of jobs and work organisation

differ. Alternatives which were seen to exist are outlined. The approach adopted for developing the "best fit" is described.

Case 1.—Designing jobs and work organisation on new manufacturing facilities in a biscuit factory

Since the early seventies there has been considerable development within the company of employee involvement in the decision-making process. In addition to the formation of consultative machinery at several levels, including factory-wide and departments within the factory, there is also experience of a participative approach to the development of work systems. The latter extended over eighteen months prior to the commissioning of the first new plant. This developmental activity focused upon investigating the impact of job redesign and work organisation change upon the motivation of employees and their job satisfaction [see 5]. An objective was to develop knowledge and expertise in job design change; evolving principles for job and work organisation design were to be applied on new facilities. Any novel work organisation tested on the first plant commissioned will provide information for the design of subsequent plants.

The initial exercises in job design and work organisation change were developed by project groups consisting of local management, trade union and employee representatives with assistance from the industrial engineering department. A demonstration project was seen as a "useful learning experience" with a fair degree of success and support for further developments coming particularly from senior stewards in the union representing operatives as well as those involved in earlier projects.

Discussions were held between management and trade union representatives during the development of the new facilities. Operatives in the department were briefed at various stages prior to and during construction. The design of the manufacturing facilities, however, was the responsibility of the project engineering department. The design process requires the integration of mechanical and electrical engineering design, biscuit technology and packaging technology. Two U.K. manufacturers supply the main part of the equipment i.e. bulk handling facilities, mixing, forming, baking and chocolate enrobing. The choice of manufacturers for handling, wrapping and packing machinery is less restricted. Innovation in the operations technology appears to be stimulated by the equipment supplier rather than the biscuit manufacturer. Much of the equipment new to this factory had been installed in other organisations previously.

In order to stimulate discussion about appropriate job and work organisation design a Company Industrial Engineer, with a specific responsibility for examining such aspects, and an outside researcher undertook a study comparing the first plant commissioned with the plant being replaced. The impact of technological changes upon job content was assessed using a scheme described in [2][2]. Each task was rated on 25 dimensions each on a nine point scale. The resulting scores were computed to provide attribute scores corresponding to the ten job attributes listed earlier. It was found that generally those jobs scoring at the mid-point or above on many of the job attributes on the earlier plant were further enhanced as a result of the technical changes. One job was enlarged as a result more of organisational changes than technical changes because of the integration of two former jobs. There was an increase in discretion/responsibility and still demands for many jobs. As a direct result of technical changes one job, i.e. wrapping machine operator, scored significantly higher on most attributes than the six jobs eliminated due mainly to the need for the application of new skills requiring an understanding of information relating to the control of the process, and a command of procedures which enable the worker to respond in varying conditions. This employee will have a greater freedom of movement than those formerly performing related tasks by manual methods. This greater freedom of movement will facilitate social interaction, compensating for the reduced opportunities available whilst machine operating (noise and relative isolation being restricting factors).

Some jobs, however, remained largely unaffected particularly in later stages of the process, i.e. packing into cartons and cardboard outers, where there was no technical change.

[2] Further details of this study are available from the author.

In fact it appeared that sources of work variety available for those in the low-scoring jobs on older plants including limited rotation between work stations, change resulting from redeployment when production was interrupted, or changes in product resulting in different work methods, were no longer available to those now performing similar tasks. It was concluded that the changes had led to further "impoverishment" of those jobs previously low in terms of desired attributes.

Industrial engineering accepts the principles stated earlier for designing jobs. An attempt is being made to develop a "fit" between job and personal needs as well as providing the opportunity for personal development through the job. It is realised that the expectations and capabilities of individuals vary, therefore the system must be designed with flexibility if individuals are to be permitted to achieve need fulfillment by different means. Further, since the development of team spirit is an objective, the make-up and tasks of the work group must be considered. The introduction of new manufacturing facilities has made possible a re-examination of jobs and work organisation. Clearly the technology imposes constraints upon the nature of the jobs since many tasks are defined by the mechanical processes selected.

Alternative methods of work organisation have been identified, however, which could offer opportunities to develop jobs at the later section of the plant with enhanced motivational prospects. Employees could feasibly be more involved in decisions relating to both intra-system organisation and system regulation. It is not intended however that a radically different job and work organisation design be imposed on operatives but rather the jobs would be developed at a rate meeting the personal needs of the job-holders.

Certain factors limit the extent to which certain changes will be acceptable. Biscuit manufacture has existed in the factory for over 25 years and there are well established traditions and custom and practice. The factory organisation separates the functions of quality control and maintenance from the biscuit department management. Within the biscuit department factory records are separated from line management. Without the acceptance that a complete review of the organisation structure may be necessary, it appears unlikely that changes in the role of the work group will result.

In the current strategy for the development of the job design "fit" there are three main elements. This approach has resulted from experiences gained during the earlier developmental projects. The steps taken are as follows:

1. *Training* for managers, supervisors and shop stewards which included discussion of the principles of job and work organisation design, an introduction to joint problem-solving and discussion about the overall development of employee participation. Seminars for all employees on the plant covering the above in less detail.

2. *Communication meetings* to be held regularly by local management. Items of direct relevance to the plant will be discussed as well as information provided on broader issues by both management and trade union.

3. A *joint problem-solving group* has been established. "Expert" assistance is available to the group of both a technical nature and regarding developing effective problem-solving skills.

The work group is being encouraged to resolve its own problems with less reliance upon supervision and management. In the case of problems needing immediate solution, management is encouraged to be supportive and adopt a "participative style". Solutions to longer-term problems, where appropriate, will be developed by joint problem-solving groups.

Case 2.—Developing an assembly and test system for manufacturing electronic equipment

The company involved in this second case was established to manufacture two electronic products for three customers. The factory to be used for the operation is situated near London in an area where there is a tradition of electrical and electronic assembly. It is anticipated that the demand for the two product ranges will develop fairly rapidly as manufacture is transferred from an American parent company. The eventual labour force involved in the assembly and test is expected to be approximately 300 with technical services and management also employing about 300. At the

time of the study most managerial appointments had been made, some technical personnel had been engaged in product development for approximately 12 months and others had been involved in designing manufacturing methods for several months. The four initial supervisory appointments had been made. The build-up of the labour force was due to commence one month following this study.

Information about the product is available from the American parent but further product development is underway in the U.K. prior to commencement of manufacture. Information about assembly and test methods is also available but again this is not necessarily applicable for the modified product and in situations where the opportunity to develop new manufacturing facilities exists. It is anticipated, particularly initially, that product developments will have to be incorporated into both part-complete and assembled equipment. This will lead to increased work loads due to rework.

The company has a declared policy for all employees of equal status and equal opportunities for personal development. The approach to managing the operation to be adopted is one of full consultation on matters of significance.

A two-day workshop was set up with the objective of designing a work system utilising available knowledge about "best practice" so that jobs meet both individual and organisation needs. This was attended by those personnel with responsibility for designing work systems, i.e. manufacturing manager, 2 supervisors, 2 production engineers, 2 manufacturing engineers, 2 quality control engineers, personel manager and training manager. Two outside consultants developed the programme, made theoretical inputs and organised small group working.

In order to establish objectives for the work system design relating to the needs of both employees and the organisation, during an early part of the workshop each participant was asked to:

a. List from present and past experience those aspects which have been the source of greatest difficulty(ies) when working with production systems.

b. Complete short questionnaires asking for views about the "typical production employee" and "ideal job characteristics and organisation structure" for the production system.

Previous experiences identified several problems to be avoided and issues to be considered in the present design process. The problems and issues raised covered a range of aspects of work system design. The exercise served to reinforce the need to consider human aspects in the design process and means for developing good working relationships, flexibility, appropriate evaluation and rewards, and opportunities for development.

The second stage in this small group activity involved each participant completing a questionnaire concerned with "characteristics of the typical production employee". A second questionnaire "the ideal job characteristics and organisation structure" was completed. The results were collated by the consultant within the small groups and immediate feedback presented. Differences in views and their significance when designing work systems were discussed by participants. This revealed that some participants held views constantly more extreme than their colleagues as well as views about job design which differed considerably when considering their own job compared with those at shop-floor level. The questions asked, the mean scores for all participants and the range of scores are presented in Tables 2 and 3.

During the next period information was presented to the group about the psychological requirements of jobs, job attributes and principles for work group design as described earlier. Participants then developed a flow diagram and explanation of the sequence of manufacture as well as a statement outlining the current stage in development of the work system. The work of individuals in the design team was integrated and the relationships with other sections established. The small groups then proceeded to identify feasible alternative methods of arranging the tasks and organising the work. The alternatives identified were evaluated using the list of job attributes.

One group concerned with the complete assembly and testing of a large, complex equipment identified five alternative patterns of work organisation:

1. operator at fixed work station, part-complete

TABLE 2. THE CHARACTERISTICS OF THE TYPICAL PRODUCTION EMPLOYEE

				RANGE
A.	Leaves other people to make most of the decisions on things which affect them at work.	1 X 7	Will protest if they are not consulted on all matters which affect them at work.	2–5
B.	Capable of handling only a limited range of tasks in their job	1 X 7	Capable of doing a job involving a variety of different tasks.	2–7
C.	Not concerned about having social contact at work	1 X 7	Regards opportunities for social contact at work as important.	4–6
D.	Can tolerate boring work.	1 X 7	Demands interesting work.	2–6
E.	Work best if the pace of the work is outside control	1 X 7	Has complete control over the pace of work.	4–7
F.	Needs or wants to have a well-defined job (area of operation) which he/she sticks to most of the time.	1 X 7	Works well and enjoys working in a job (area of operation) which is not clearly defined.	2–6
G.	Needs to be told what to do next and how to do it.	1 X 7	Can organize the sequence of work and choose the best methods.	2–6
H.	Unable to undertake responsibility for decisions & unable to take initiative.	1 X 7	Able to undertake responsibility for decisions and able to take initiative.	2–7
I.	Has low level of skill and/or knowledge (expertise)	1 X 7	Has a high level of skill and/or knowledge (expertise)	3–7

Based on: James C. Taylor. *A Report of Preliminary Findings from the 1976 'Working Organisation Study' Pilot Study* (Centre for Quality of Working Life. UCLA Reprints. 1976).

assemblies brought to work-station, limited work rotation.

2. operator moving with individual assembly through total process

3. 4 work groups—sub-assembly; chassis assembly; deck assembly;

4. 3 work groups—sub-assembly and chassis; sub-assembly and deck; final assembly and test.

5. Work groups each building a complete unit.

Each operator would be multi-skilled. When working in a group, operators would each be trained in all the tasks. Testing the product at various stages would be the responsibility of assemblers. Three alternatives were evaluated against the ten job attributes by means of a three point scale and scores summed to give an overall assessment of each solution in terms of their motivational potential. The most favourable solution was found to be number 5 with the least number 1. This group decided that the next step, to be undertaken outside the workshop, would be to carry out a more thorough exercise developing the work organisations more fully and evaluating them on both quantifiable and non-quantifiable criteria. They are to report to colleagues three months following the workshop.

TABLE 3. THE IDEAL JOB CHARACTERISTICS AND ORGANIZATION STRUCTURE

			RANGE
A.	Decisions on what is to be done and how it is to be done should be left entirely to management.	Decisions should be arrived at through group discussions involving all employees.	2-7
B.	Job methods should be carefully defined by systems and procedures specialists, management services, or supervision.	The development of job methods should be left to the group and individual doing the job.	1-5
C.	The most important motivators should be financial, e.g. high earnings and cash bonuses.	The most important motivators should be non-financial, e.g. work challenge, opportunity for team work.	2-7
D.	There should be close supervision, tight controls, and well-maintained discipline.	There should be loose supervision, few controls and a reliance on employee self-discipline.	1-6
E.	Groups and individuals should be given the specific information they need to do the job but no more.	Everyone should have access to all information which they regard as relevant to their work.	5-7
F.	Jobs should be clearly defined, structured, and stable.	Jobs should be flexible and permit group problem solving.	2-6
G.	Targets should be set by supervision and monitored by supervision.	Targets should be left to the employee groups to set and monitor.	2-5
H.	There should be a clear hierarchy of authority, with the person at the top carrying ultimate responsibility for all aspects of work.	There should be a delegation of authority and responsibility to those doing the job regardless of formal title and status.	2-6

Based on: James C. Taylor, *A Report of Preliminary Findings from the 1976 'Work Organization Study' Pilot Study* (Centre for Quality of Working Life, UCLA Reprints, 1976).

CONCLUSIONS

The work system designer is required to produce a system capable of achieving certain objectives relating to the conversion of inputs to outputs. This requires an integration of both technical and social factors in order to maintain long-term economic viability of the operating system. The selection of appropriate work systems requires decisions which put differing emphasis upon criteria which may conflict. Trade-offs are necessary and under these conditions the designer will seek a solution which satisfices.

The quality of working life for job-holders responsible for both managing and operations within the system will be influenced by decisions made by the work system designer. The principle of minimal critical specification by the designer was proposed as a means for creating opportunities for the later development of a "fit" between the needs of individual employees and organisational objectives. The extent of opportunity for task performers to

make decisions relating to the management of the work system was seen to be a factor limiting the degree to which jobs could offer attributes believed to contribute to employee motivation.

Two case examples were described in which quality of working life issues were considered during the design process. The stage at which these issues were debated and the strategies adopted differed. In the first case, where replacement facilities were being designed, discussions and decisions about jobs and work organisation were made during and subsequent to the commissioning of the plant. Many factors influenced decisions including established custom and practice, minimisation of wage payments, traditional roles and factory organisation. Resulting opportunities to create jobs with attributes such as variety, responsibility, control and social opportunities were further restricted by technological decisions made earlier during the design process. Some decisions about operations technology were influenced by equipment availability. Those responsible for task performance, however, were to be involved in the design of jobs at this stage through a participative problem-solving approach. Additional involvement is available through the negotiation process.

In the second case, management and functional specialists considered human aspects of work-system design at an early stage in the design process. Even so, some aspects of design were already determined. The American parent company had standard manufacturing procedures and work-bench layouts. However, choices were still apparent and received consideration in the extent to which desired job attributes appeared to exist. At this stage those to be ultimately responsible for task performance were not available to be consulted about, or participate in the process of, the design of work systems. The project team had to rely upon previous experiences and available knowledge about human aspects of work system design. The approach adopted highlighted some differences in values and beliefs held by members of the design team. Discussion led to agreement about appropriate job and work group design principles to be applied but they were not able to benefit from direct interaction with task performers as a means of establishing the validity of design assumptions being made. The specific allocation of decisions of both an intra-system and a regulatory nature was deferred to be discussed between task performers and local magagement at a later stage. Certain decisions affecting the work organisation, however, could not be deferred. These tended to relate to the boundaries of the system and specifically resource availability.

Decisions made during the various stages of work system design have an impact upon the quality of working life of employees. At each stage decisions made limit choices available at later stages. Without a radical review of the organisation the benefits to be derived in terms of both productivity and job satisfaction may be missed. The work system designer, if these opportunities are to be taken, must consider not only the technical aspects of system but also the design of the organisation, so as to meet organisational objectives and individual needs.

Vol. 6
1978

ACKNOWLEDGEMENTS

The assistance of Jenny Blake and Don Helsby in collecting case material is gratefully acknowledged.

REFERENCES

1. BIRCHALL DW (1975) *Job Design: A Planning Guide for Managers*. Gower Press, Epping, Essex, UK.
2. CARNALL CA, BIRCHALL DW & WILD R (1976) The design of jobs—an outline strategy for diagnosis and change. *Mgmt Serv.* **20**(6), 48–51.
3. CHERNS A (1977) Can behavioural science help design organisations? *Org. Dynamics*, **6**(1), 44–74.
4. EMERY F & THORSRUDE (1976) *Democracy at Work*. Mastimus Nijhoff Social Sciences Division, Leiden, Holland.
5. HELSBY DD, BIRCHALL DW & CARNALL CA (1977) A role for the industrial engineer in a job redesign project. *Mgmt Serv.* **21**(11), 16–19.
6. SCHUMACHER C (1976) *Principles of Work Structuring*. British Steel Corporation, London.

ADDRESS FOR CORRESPONDENCE: *David W Birchall Esq., Henley, The Administrative Staff College, Greenlands, Henley-on-Thames, Oxfordshire RG9 3AU, UK.*

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 443–448
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0443/$02.00/0

Some Practical Considerations on Multi-Server Queues with Multiple Poisson Arrivals

GEORGE P COSMETATOS

Imperial College, London

(*Received May 1978*)

A simple approximate formula is presented and evaluated in this paper for the steady-state average number of customers in the $M^{(x)}/G/r$ queueing system. The derivation of the formula is based on a heuristic argument. Comparisons between approximate results and the ones obtained analytically or by extensive simulation experiments seem to suggest that the degree of approximation achieved is satisfactory for most practical purposes.

INTRODUCTION

QUEUEING systems where customers arrive in groups and/or get served in batches are quite frequent in practice. However, models of such systems are seldom amenable to analytic methods leading to directly applicable results, and in the absence of approximation techniques or solutions the practitioner may have no alternative than to resort to simulation. Yet this approach has its drawbacks: it is time consuming and not always easy to design, analyse or evaluate.

A useful model which does not seem to lend itself to easy mathematical treatment is the one described as $M^{(x)}/G/r{:}(\infty/GD)$ in Lee's [7] notation. It is assumed in this model that:

(a) Customers arrive in groups (batches) of size X, where X is an integer random variable whose probability distribution is $Pr\{X = x\} = c_x (1 \leq x < \infty)$ with mean $E(x)$ and variance var(x).
(b) Arrivals of batches of size x form a Poisson process with mean λ_x; the average rate of arrival of customers is $\lambda E(x)$, where

$$\lambda = \sum_{x=1}^{r} \lambda_x.$$

(c) There are r servers in parallel; each server has an independently and identically distributed service-time distribution $B(t)$; the average service time is μ^{-1} and the coefficient of variation of service times v.
(d) Arriving customers join a single queue and are served individually.
(e) The system is in statistical equilibrium, i.e. the traffic intensity $\rho = \lambda E(x)/r\mu < 1$.

Let $L_x(r,v)$ denote the average number of customers in the system (those in the queue and the ones being served). In this paper an approximate formula for $L_x(r,v)$ is developed and the degree of approximation achieved under various conditions investigated. The derivation of the formula is based on a heuristic argument whereby a reformulation of the expression for $L_x(1,v)$ is extended to the multi-server queue. From a computational viewpoint, the formula is very simple to apply and the relative percentage errors incurred seem to be quite small.

THE $M^{(x)}/G/1$ QUEUE

Saaty [12] derives the following expression for the probability generating function (*pgf*) of the number of customers in the system:

$$P(z) = \frac{(1-\rho)(1-z)K(z)}{K(z)-z}, \qquad (1)$$

where $K(z) = B^*[\lambda - \lambda C(z)]$, with $B^*(s)$ and $C(z)$ denoting respectively the Laplace–Stieltjes transform of the service-time distribution and the *pgf* of the variable X.

The average number of customers in the system can now be obtained by differentiating (1) with respect to z and setting $z = 1$ after using L'Hôpital's rule twice:

$$L_x(1,v) = \frac{\rho}{2(1-\rho)}[E(x) + 1 - \rho(1 - v^2) + \mathrm{var}(x)/E(x)]. \quad (2)$$

This result indicates that by removing the *pgf* curtain in (1) the expected congestion in the system becomes a function of four parameters only and does not depend on the shapes of either the service-time or the batch-size distributions. It is interesting to note that the expression in (2) can be re-arranged in three different ways:

(a) $$L_x(1,v) = L_1(1,v) + \frac{\rho}{2(1-\rho)}[E(x) - 1] + \frac{\rho}{2(1-\rho)} \cdot \frac{\mathrm{var}(x)}{E(x)}. \quad (2a)$$

Thus $L_x(1,v)$ can be viewed as a sum of three distinct terms: the first term corresponds to the average number of customers in the standard $M/G/1$ system with traffic intensity ρ and coefficient of variation v, given by the Pollaczek–Khintchine formula:

$$L_1(1,v) = \frac{\rho[2 - \rho(1 - v^2)]}{2(1-\rho)} \quad (3)$$

or

$$L_1(1,v) = v^2 L_1(1,1) + (1 - v^2) L_1(1,0). \quad (3a)$$

where $L_1(1,1)$ and $L_1(1,0)$ correspond to the systems $M/M/1$ and $M/D/1$ respectively. The second reflects the additional congestion caused by the $E(x) - 1$, on average, members of each batch who join the system upon arrival but cannot be immediately served. Finally, the third term gives the effect on $L_x(1,v)$ of the variability in the batch size. The relative contribution of each of these factors to the system's overall congestion can easily be assessed, given the simplicity of the expression in (2a).

(b) $$L_x(1,v) = v^2 L_x(1,1) + (1 - v^2) L_x(1,0) \quad (2b)$$

where $L_x(1,1)$ and $L_x(1,0)$ correspond to the systems with exponential ($v = 1$) and constant ($v = 0$) service times respectively. Thus, (2b) is a generalization of (3a) and clearly suggests that the Pollaczek–Khintchine formula also applies to the more general system considered in this section.

(c) $$L_x(1,v) - L_1(1,v) = L_x(1,1) - L_1(1,1)$$

or

$$L_x(1,v) - L_1(1,v) = L_x(1,0) - L_1(1,0). \quad (2c)$$

These two expressions point to the conclusion (also apparent from (2a)) that the additional congestion which is created in the system by the multiple nature of arrivals is independent of the coefficient of variation of service times.

AN APPROXIMATION FOR $L_x(r,v)$

It is now postulated that the conclusion drawn in the previous section can be generalised to provide the two approximate formulae:

$$L_x(r,v) \simeq L_x(r,1) - [L_1(r,1) - L_1(r,v)] \quad (4)$$

and

$$L_x(r,v) \simeq L_x(r,0) + [L_1(r,v) - L_1(r,0)] \quad (5)$$

for $L_x(r,v)$, where all L's are calculated for the same traffic intensity ρ. One would expect (4) to perform better than (5) if v is close to 1 and (5) to perform better than (4) if v is very small. However, in the absence of analytical results for $L_x(r,0)$ the idea of using (5) or some linear combination of (4) and (5) as an approximation for $L_x(r,v)$ has to be abandoned.

With reference to formula (4), $L_1(r,1)$ and $L_1(r,v)$ denote the average number of customers in the systems $M/M/r$ and $M/G/r$ respectively. Although analytical results for $L_1(r,v)$ are generally not available, the approximation

$$L_1(r,v) \simeq v^2 L_1(r,1) + (1 - v^2) L_1(r,0)$$

established in [2] in conjunction with the explicit formula suggested in reference [1] for the approximate evaluation of $L_1(r,0)$ yield:

$$L_1(r,1) - L_1(r,v) \simeq \rho r(1 - v^2)\left[\frac{1}{2} - \frac{(1-\rho)(r-1)(\sqrt{4+5r} - 2)}{32\rho r}\right]\mu W(r,1) \quad (6)$$

where $W(r,1)$ is the average queueing (excluding service) time of an arriving customer in the standard $M/M/r$ queue calculated for traffic intensity ρ. Tabulated results for $\mu W(r,1)$ are widely available (see, for example, [10]) and, hence, the evaluation of the right-hand side in (6) presents no computational difficulties; the relative percentage errors on the other hand which are incurred when using the approximation in (6) are negligible, not exceeding 2% in absolute value over a wide range of values for ρ, r and v.

$L_x(r,1)$ in formula (4) is the average number of customers in the system $M^{(x)}/M/r$. Its calculation is discussed in the next section.

THE $M^{(x)}/M/r$ QUEUE

This queueing model was first studied by Kabak [6], who proposed a general numerical approach to the solution of the steady-state probabilities. The analysis by Kabak assumes that the batch size is constant. This assumption can, however, be relaxed (see, for example, Gross and Harris [5]) in which case the steady-state difference equations for the probabilities p_n of n in the system are:

$$0 = -\lambda p_0 + \mu p_1 \qquad (n = 0) \qquad (7a)$$

$$0 = -(\lambda + n\mu)p_n + (n+1)\mu p_{n+1} + \lambda \sum_{x=1}^{n} p_{n-x}c_x \qquad (1 \le n < r) \qquad (7b)$$

$$0 = -(\lambda + r\mu)p_n + r\mu p_{n+1} + \lambda \sum_{x=1}^{n} p_{n-x}c_x \qquad (r \le n) \qquad (7c)$$

Using a generating-function approach to solve the above system of equations one finally obtains:

$$L_x(r,1) = \frac{\rho}{2(1-\rho)}\left\{\frac{2}{\rho}\left[\sum_{n=1}^{r-1} np_n - \frac{1}{r}\sum_{n=1}^{r-1} n^2 p_n\right] + 1 + E(x) + \mathrm{var}(x)/E(x)\right\} \qquad (8)$$

where

$$r\sum_{n=0}^{r-1} p_n - \sum_{n=1}^{r-1} np_n = r(1-\rho)$$

is the normalising equation.

In general, the probabilities p_n $(n = 1, \ldots, r-1)$ cannot be expressed in closed form. If r is small, an iterative method of solving the steady-state equations is possible; otherwise one has to resort to numerical methods [4]. However, it is interesting to note that if the probability of a batch having a size smaller than $r-1$ is negligible, and this is always true for $r = 2$, then $c_1, c_2, \ldots, c_{r-2} \simeq 0$ and (7b) is simplified to:

$$0 = -(\lambda + n\mu)p_n + (n+1)\mu p_{n+1} \quad (1 \le n < r-1)$$

$$0 = -[\lambda + (r-1)\mu]p_{r-1} + r\mu p_r + \lambda p_0 c_{r-1} \quad (n = r-1).$$

In this case it can be easily shown that

$$p_n = \frac{r!(1-\rho)\prod_{j=0}^{n-1}[j + \rho r/E(x)]}{n!\prod_{j=2}^{r}[j + \rho r/E(x)]} \qquad (0 \le n \le r-1)$$

where the empty product is equal to 1, and the expression in (8) is greatly simplified:

$$L_x(r,1) = \frac{\rho r(r-1)}{\rho r + 2E(x)} + \frac{\rho}{2(1-\rho)} \times [E(x) + 1 + \mathrm{var}(x)/E(x)]. \qquad (9)$$

If the probability of a batch having a size smaller than $r-1$ is not negligible, one can still use formula (9) as an approximation for $L_x(r,1)$. Suppose, for example, that the batch size has the geometric distribution

$$c_x = (1-a)a^{x-1} \quad (0 < a < 1)$$

with $E(x) = 1/(1-a)$ and $\mathrm{var}(x) = a/(1-a)^2$. Then $\mathrm{var}(x)/E(x) = a/(1-a) = E(x) - 1$, and

TABLE 1. RELATIVE PERCENTAGE ERRORS INCURRED IF FORMULA (10) IS USED FOR THE APPROXIMATE EVALUATION OF $L_x(r,1)$: THE BATCH-SIZE DISTRIBUTION IS ASSUMED TO BE GEOMETRIC

r	ρ \ $E(x)$	2	3	4	5	10	20
3	0.5	2.2	0.9	0.5	0.3	0.0	0.0
	0.6	1.6	0.7	0.4	0.2	0.0	0.0
	0.7	1.1	0.5	0.2	0.1	0.0	0.0
	0.8	0.7	0.3	0.2	0.1	0.0	0.0
	0.9	0.3	0.1	0.1	0.0	0.0	0.0
4	0.5	5.9	2.7	1.5	0.9	0.2	0.0
	0.6	4.3	2.0	1.1	0.7	0.1	0.0
	0.7	2.9	1.4	0.8	0.5	0.1	0.0
	0.8	1.8	0.9	0.5	0.3	0.0	0.0
	0.9	0.8	0.4	0.2	0.1	0.0	0.0
5	0.5	10.2	5.2	3.0	1.8	0.3	0.1
	0.6	7.5	3.9	2.2	1.4	0.3	0.0
	0.7	5.2	2.7	1.5	1.0	0.2	0.0
	0.8	3.2	1.6	1.0	0.6	0.1	0.0
	0.9	1.5	0.8	0.4	0.3	0.1	0.0

(9) becomes

$$L_x(r,1) \simeq \frac{\rho r(r-1)}{\rho r + 2E(x)} + \frac{\rho}{1-\rho} E(x). \qquad (10)$$

The relative percentage errors incurred if (10) is used for the approximate evaluation of $L_x(r,1)$ are summarised in Table 1 for $r = 3$, 4 and 5, $\rho = 0.5$, 0.6, 0.7, 0.8 and 0.9 and $E(x) = 2, 3, 5, 10, 20$. Provided that the average batch size is not too small relative to the number of servers the degree of approximation achieved seems to be reasonably good.

EVALUATION OF THE APPROXIMATE FORMULA IN (4)

Following the discussion in the last two sections, the calculation of the right-hand side in (4), using (6) and (9), is straightforward. However, in order to test the degree of approximation achieved, analytical or simulated values for $L_x(r,v)$ over a wide range of values of the parameters involved are clearly required. The cases that were examined are summarised in Table 2.

In cases 1 and 2 it is assumed that service times have the Erlang distribution

$$dB(t) = \frac{(k\mu)^k t^{k-1} \exp(-k\mu t)}{(k-1)!}\, dt$$

with mean $1/\mu$, variance $1/k\mu^2$ and coefficient of variation $v = 1/\sqrt{k}$.

This distribution is identical to the distribution of the sum of k independent random variables having an exponential distribution with mean $1/k\mu$, and it is this property which provides the key to the analysis of $M^{(x)}/E_k/r$ systems. Following the methodology developed by Poyntz and Jackson [11] the steady-state difference equations describing the system under consideration were solved for the values of r and v^2 indicated in Table 2 and analytical results for $L_x(r,v)$ were obtained over the whole range of values for ρ and $E(x)$. In cases 3, 4, 5 and 6 simulation studies were conducted using the SIMON [8] language and the Imperial College CDC 6400 computer. Eighteen models were run and the assumptions regarding the service-time distribution and the values of the parameters involved are shown in Table 2. In each run 40,000 arriving batches were simulated (the requirements in computer time being equivalent to approximately $6\frac{1}{2}$ min on a CDC 6400 computer) and the average number of customers in the system was calculated from approximately 200,000 observations taken at random points in time.

A comparison between the simulated results in cases 4, 5, 6 and the analytical ones in cases 1 and 2 seems to point to the tentative conclusion that, as in the single server system and indeed other multi server systems with Poisson arrivals and general service times [2, 3, 9] $L_x(r,v)$ does not depend on the shape of the service-time distribution but only on its mean and variance: The analytical values for $L_x(r,v)$ obtained in cases 1 and 2 were found to lie well within the 95% confidence limits on the simulated values in cases 4, 5 and 6.

The comparison between the approximate values for $L_x(r,v)$, as calculated from formulae (4), (6) and (9) and the corresponding analytical

TABLE 2. CALCULATION OF $L_x(r,v)$: CASES EXAMINED

Case	D tion of service times	r	v^2	ρ	D tion of the batch size	Average batch size	Calculation of $L_x(r,v)$
1	Erlang	2	$\frac{1}{2}, \frac{1}{3}$	0 1	Determin. Geometric	2, 3, 4, 5 10, 20	Analytically
2	Erlang	3, 4	$\frac{1}{2}$	0 1	Determin. Geometric	2, 3, 4, 5 10, 20	Analytically
3	Determin.	2, 4	0	0.5, 0.6 0.7, 0.8	Determin.	5	By simulation
4	Triangular	2, 4	$\frac{1}{2}$	0.6, 0.8	Determin.	5	By simulation
5	Rectangular	2	$\frac{1}{3}$	0.6, 0.8	Determin.	2, 5	By simulation
6	Discrete	2, 4	$\frac{1}{2}$	0.6	Determin.	5	By simulation

TABLE 3. RELATIVE PERCENTAGE ERRORS INCURRED IF FORMULAE (4), (6) AND (9) ARE USED FOR THE EVALUATION OF $L_x(r,v)$ IN THE $M^{(x)}/E_2/r$ QUEUE: THE BATCH SIZE IS ASSUMED TO BE CONSTANT

ρ	r \ $E(x)$	2	3	4	5	10	20
0.5	2	0.5	−0.7	−1.1	−1.2	−1.0	−0.6
	3	−0.2	−0.0	−1.1	−1.6	−1.9	−1.3
	4	2.8*	−0.6	−0.6	−1.4	−2.5	−2.0
0.6	2	0.4	−0.5	−0.8	−0.8	−0.7	−0.5
	3	−0.1	0.0	−0.7	−1.2	−1.4	−1.0
	4	2.0*	−0.5	−0.4	−1.0	−1.9	−1.5
0.7	2	0.3	−0.3	−0.5	−0.6	−0.5	−0.3
	3	−0.1	0.1	−0.5	−0.8	−1.0	−0.8
	4	1.4*	−0.3	−0.2	−0.7	−1.3	−1.0
0.8	2	0.2	−0.2	−0.3	−0.3	−0.3	−0.2
	3	−0.0	0.1	−0.3	−0.5	−0.6	−0.4
	4	0.9*	−0.2	−0.1	−0.4	−0.8	−0.6
0.9	2	0.1	−0.1	−0.1	−0.1	−0.1	−0.1
	3	−0.0	0.0	−0.1	−0.2	−0.3	−0.2
	4	0.4*	−0.1	−0.0	−0.2	−0.4	−0.3

* If the approximation in (9) is replaced by the analytical values of $L_x(r,1)$, the relative percentage errors corresponding to $\rho = 0.5, 0.6, 0.7, 0.8, 0.9$ become −0.1, −0.1, 0.0, 0.0 and 0.0 respectively.

or simulated ones can be summarised as follows:

(a) For a given batch-size distribution and a given value of v^2, $L_x(r,v)$ can be approximated with a relative percentage error e which:

is positive or negative depending on the values of the parameters involved and tends to 0 as $\rho \to 1$, $v \to 1$ and/or $E(x) \to \infty$;

is not sensitive to the values of ρ, r or $E(x)$;

is generally quite small in absolute value and cannot, therefore, be easily identified if analytical results are not available.

As an indication of the magnitude of e, Table 3 gives the relative percentage errors incurred if formulae (4), (6) and (9) are used for the evaluation of $L_x(r,v)$ in the $M^{(x)}/E_2/r$ queue for constant batch sizes equal to 2, 3, 4, 5, 10,

TABLE 4. ANALYTICAL, SIMULATED AND APPROXIMATE RESULTS IN THE $M^{(5)}/E_k/2$ QUEUE

ρ		0.5				0.6			
Measures of interest	$v^2 = 1/k$	1	$\frac{1}{2}$	$\frac{1}{3}$	0	1	$\frac{1}{2}$	$\frac{1}{3}$	0
Analytical value		3.09	3.05	3.04	—	4.61	4.48	4.44	—
Lower 95% confidence limit		—	—	—	3.02	—	—	—	4.32
Simulated value		—	—	—	3.04	—	—	—	4.39
Upper 95% confidence limit		—	—	—	3.06	—	—	—	4.45
Approximate value		3.09	3.01	2.99	2.93	4.61	4.44	4.39	4.28
Relative percentage error		0	−1.2	−1.7	−3.5	0	−0.8	−1.2	−2.4

ρ		0.7				0.8			
Measures of interest	$v^2 = 1/k$	1	$\frac{1}{2}$	$\frac{1}{3}$	0	1	$\frac{1}{2}$	$\frac{1}{3}$	0
Analytical value		7.12	6.83	6.74	—	12.14	11.48	11.26	—
Lower 95% confidence limit		—	—	—	6.32	—	—	—	9.94
Simulated value		—	—	—	6.56	—	—	—	10.71
Upper 95% confidence limit		—	—	—	6.80	—	—	—	11.48
Approximate value		7.12	6.79	6.68	6.47	12.14	11.44	11.20	10.73
Relative percentage error		0	−0.6	−0.8	−1.4	0	−0.3	−0.5	0.2

20, $r = 2, 3, 4$ and $\rho = 0.5, 0.6, 0.7, 0.8, 0.9$.

(b) For a given batch-size distribution and given values of ρ, r and $E(x)$, the degree of approximation achieved in evaluating $L_x(r,v)$ does not seem to deteriorate markedly as v approaches 0 (see, for example, Table 4). Judging from the cases $\rho = 0.5$ and $\rho = 0.6$, where the 95% confidence limits on the simulated values are fairly tight, the relative percentage errors incurred are negative but reasonably small in absolute value. On the basis of the results presented in Table 4 one could also argue that formula (4) may give more reliable estimates for $L_x(r,v)$ than a carefully designed simulation study, if the value of ρ is high (higher than, say, 0.7).

CONCLUSIONS

Formula (4) in conjunction with (6) seems to provide a simple analytical tool for calculating the steady-state average number of customers in a multi-server queueing system with Poisson batch arrivals and general service times. If the probability of a batch size not exceeding $r - 2$ is small, the explicit approximation in (9) simplifies the evaluation of (4) further. The degree of approximation achieved is reasonably good and insensitive to the values of the various parameters involved. If interested in redesigning an already existing system which the model discussed in this paper describes, the relative percentage errors incurred are likely to be very small.

REFERENCES

1. COSMETATOS GP (1975) Approximate explicit formulae for the average queueing time in the processes $(M/D/r)$ and $(D/M/r)$. *INFOR* **13**(3), 328–331.
2. COSMETATOS GP (1976) Some approximate equilibrium results for the multi-server queue $(M/G/r)$. *Op. Res. Q* **27**(3i), 615–620.
3. COSMETATOS GP (1977) Cobham's model on non preemptive multi-server queueing systems—a heuristic method for its generalisation. *EJOR* **1**(4), 262–264.
4. CROMIE MV (1974) Measures of efficiency for the bulk-service queue $M/M^{(x)}/c/\infty$ and the bulk-arrival queue $M^{(x)}/M/c/\infty$ with computing algorithms, tables and charts—and steady-state probability equations for special cases of the queue with bulk-arrival and bulk-service $M^{(x)}/M^{(y)}/c/\infty$ (in two vol.). MSc Thesis, Dept of Mathematics, Royal Military College of Canada, Kingston, Canada.
5. GROSS D & HARRIS CM (1974) *Fundamentals of Queueing Theory*. John Wiley, New York.
6. KABAK IW (1968) Blocking and delays in the $M^{(n)}/M/c$ bulk queueing systems. *Op. Res.* **16**(4), 830–840.
7. LEE AM (1966) *Applied Queueing Theory*. St Martin's Press, Montreal, Canada.
8. MATHEWSON SC (1978) *A DRAFT/SIMON Manual*. Dept of Management Science, Imperial College, London.
9. NAGACHANDRAN S (1976) The economic assignment of men to machines. MPhil Thesis, Dept of Management Science, Imperial College, London.
10. PAGE E (1972) *Queueing Theory in OR*. Butterworths, London.
11. POYNTZ CD & JACKSON RRP (1973) The steady-state solution for the queueing process $E_k/E_m/r$. *Op. Res. Q* **24**(4), 615–625.
12. SAATY TL (1961) *Elements of Queueing Theory*. McGraw-Hill, New York.

ADDRESS FOR CORRESPONDENCE: *Dr GP Cosmetatos, Department of Management Science, Imperial College, Exhibition Road, London SW7 2BX, UK.*

Vol. 6
1978

MEMORANDA

. 6
78

BLANK
PAGE

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 451–454
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0451$02.00/0

Factor Analysis—A Method for Reducing the Number of Criteria in a Multiple-Criteria Model

GIVEN a multiple-criteria problem with a large number of criteria, a method for reducing the number of criteria while still retaining a sufficient amount of available information is described in this paper. Using fewer criteria implies that the multiple-criteria model will be easier to use and that the necessary estimates will be more reliable and be easier to obtain. The application of this method to a real-life problem is described and the results of this study are discussed.

1. INTRODUCTION

When dealing with multiple-criteria problems the decision-maker usually says that he considers many criteria, although often only a few are essential. For example, Turban and Metersky [20] have found that normally only 4 to 7 criteria are important; they continue: "working with few measures of effectiveness from the beginning increases the chance of success, since there will be less spread in answering the questionnaires as well as better response" (p. B-827). Other reasons for using few criteria include lower cost and greater ease in estimating and using the multiple-criteria model; see also [11, 13, 14 and 21] for corroboration.

Thus it is important in all multiple-criteria work to try to reduce the proposed number of criteria. In Section 2 we will present some of the approaches suggested in literature. Section 3 is devoted to factor analysis, a technique for compressing data that has not been used previously to help decision-makers solve multiple-criteria problems. The technique has been tested on a real-life problem, which is briefly presented in Section 4, and the results of the test will be found in Section 5.

2. WAYS OF REDUCING THE NUMBER OF CRITERIA

The problem of handling a large number of criteria has not been very much treated in the multiple-criteria literature. Fishburn [5] discusses three approaches to this problem. The first is to select a subset (about 6 to 10) of the criteria most important to the decision-maker and to only use these in the analysis. Another method is first to select a subset of the more important criteria and to analyze the alternatives in terms of these criteria. Then another subset of criteria is added to those initially used, and the alternatives are analyzed with respect to this larger subset of criteria. If the results of this second analysis agree with those of the first one, the decision-maker may be satisfied with the analysis. If the outcomes of the analyses differ, the decision-maker adds a new subset of criteria and repeats the analysis.

The third approach for reducing a great number of criteria suggested by Fishburn is to select any subset of criteria and analyze the alternatives with respect to these criteria. Next a new subset of criteria is selected and processed, and this process is repeated several times. If the results of the analysis of the alternatives are not dependent on the various sets of criteria, the decision-maker may use any of the subsets in his analysis.

Keeney and Nair [12] use another approach in a study of the location of nuclear power plants in the North Western states of the US. They use three general concepts to determine whether a suggested criterion will be used in the analysis or not. Their concepts are: (a) the significance of the impact in terms of one criterion in relation to impacts as measured by other criteria, (b) the variation of the impact between the alternatives in terms of one criterion, and (c) the likelihood of occurrence of significant impacts as measured by a criterion.

All these methods for reducing the number of criteria exclude in various ways a subset from the analysis. The methods do not give any information on what effect this reduction will have on the decision. Factor analysis represents a quite different approach as it retains all the original criteria but transforms them into a new smaller set of criteria (factors). As a reduction in the number of criteria (factors) is made some information is of course lost, but the decision-maker always knows how much information is lost. Thus the decision-maker can trade the loss of information against what number of criteria (factors) he wishes to maintain.

There is also another advantage in factor analysis as the created criteria may represent aspects that are more familiar to the decision-maker and thereby ease the decision-maker's stating of his preferences. These estimates will then also become more reliable. This will be illustrated by our real-life problem.

3. FACTOR ANALYSIS

Factor analysis has become the name of a variety of procedures for reducing and summarizing data. The aim of the technique is to explain observed relationships among numerous criteria in terms of simpler relationships, e.g. by forming a smaller set of hypothetical criteria that may be interpreted by available theories or knowledge. One frequent application has been the psychologists' search for dimensions (factors) underlying the manifest variables. For an overview of factor analysis and applications see [9].

The common element in these techniques is the collection of measurements of the criteria for a number of alternatives or objects. The next step is to calculate the intercorrelations of the criteria. These correlations are used to find the linear combination (a factor) of the criteria that accounts for most of the variance in the data. The next factor is the linear combination of the criteria that accounts for most of remaining variance *and* is orthogonal to the first. The process of creating factors continues until some pre-specified standard is fulfilled, at which point not all of the variance in the data is explained.

If we had wished to explain all of the variance we would have had to extract as many factors as we had criteria. Thus our desire of having fewer criteria (fac-

TABLE 1. DESCRIPTION OF THE TEN INTERCEPTION TECHNIQUE TESTS (ITT); FICTITIOUS DATA

Criteria (interception technique test)	Modern weapon	Use of counter measures	Altitude	Speed	Evasive action	Method of interception†	Visibility††
ITT1	No	No	High	High	No	LPI	VMC
ITT2	No	No	High	Medium	No	CCI	IMC
ITT3	No	No	Medium	Low	No	CCI	IMC
ITT4	No	No	Medium	Medium	No	LPI	IMC
ITT5	No	Yes	Medium	Medium	No	CCI/LPI	VMC at night
ITT6	No	No	Low	Medium	Yes	CCI	VMC
ITT7	Yes	No	Low	Medium	Yes	LPI	VMC at night
ITT8	Yes	No	Medium	Medium	No	CCI	IMC
ITT9	No	No	High	Medium	No	LPI	IMC
ITT10	No	Yes	High	Medium	Yes	CCI	VMC at night

† LPI = Lead Pursuit interception: CCI = Collision course interception.
†† VMC = Visual meterological conditions: IMC = Instrument meteorological conditions.

tors) has to be traded against a loss in the amount of explained variance.

Having extracted a small number of factors, the next problem is to label them, i.e. give some interpretation of the factors that fits our knowledge of the problem at hand. Factor analysis can help us somewhat by rotating the factor matrix. The rotation most suitable for handling the reduction of the number of criteria is the varimax approach: this is also the rotation most frequently used. It implies that the factors retained are rotated around the point of origin (still orthogonal and without altering the amount of explained variance) so as to get as few high factor loadings per factor as possible.

4. THE MULTIPLE-CRITERIA PROBLEM

The introduction of a new Planning, Programming, and Budgeting System in the Swedish defence sector implied that, among other things, the performance of all combat units had to be estimated, as costs were to be compared with performance in all decisions. Determining the capability of a combat unit is a multiple criteria problem as no single criterion can describe sufficiently the performance of such a unit.

To construct a multiple-criteria model for evaluating the interception capability of figher squadrons, a group of officers participated in a revised Delphi procedure (for details see [18]) in order to select suitable criteria. The experts agreed upon ten criteria that represented tests of various interception techniques, e.g. interception by IR-missiles or interception by guns. The panel also agreed upon other criteria, but they will not be discussed further in this article.

Let us call the criteria ITT1 to ITT10 (interception technique test). They are listed in Table 1, characterized in terms of seven variables.

Having data about the performance of 36 squadrons regarding these criteria, it was decided to attempt to reduce their number by factor analysis. Analyzing the data, we hoped to find factors that could be interpreted by the characteristics presented in Table 1. As an example, assume that one factor would get high loadings on (consist of) the ITTs 2, 3, 4, and 8. This would indicate a possibility of substituting the criterion "interception at IMC" for those four ITTs.

The advantages of such substitutions are many. For one it will be easier for the officers to assign weights to compound criteria such as "interception at IMC" than to the individuals' tests. The weights should require less time to estimate and produce more reliable estimates. The analysis may also indicate that some of the tests may be omitted, thus reducing the costs of testing the capability of the squadrons.

5. THE RESULTS[1]

As we mentioned in Section 3, the varimax rotation will be performed only on the factors retained. Consequently it is important to determine a decision-rule for how many factors to extract. One rule that has been advocated is to exclude factors with eigen values less than 1.0, as such factors may be shown to be insignificant [8]. Another rule is to extract factors until a prespecified amount of the total variance has been explained. We decided to follow this later rule and set our standard to *about* 90% of the total variance.

The rotated five-factor matrix is found in Table 2. Factor loadings (absolutely) greater than 0.75 are denoted with two asterisks, while loadings between and including 0.50 and 0.75 (absolute numbers) are marked with one asterisk. Two factors have eigenvalues greater than 1.0, and two more are very close to 1.0. The five factors selected explain 86% of the variance. Six factors would explain more than 90% of the variance, but the set of five factors was preferred in practice because a better interpretation of the factors could be given.

To ease the labeling of the factors we can look at Table 3, where we have reformulated the information from Tables 1 and 2. We find three one-criterion-factors, i.e. ITT3, ITT5, and ITT10 that we may label "low-speed/medium-altitude interception", "interception at the use of counter measures", and "interception at the use of counter-measures and at evasive action", respectively. We find that ITT1, ITT2, ITT4, and ITT9 are grouped together into something like a "high/medium-altitude-and-speed" factor, as it is dominated by the high altitude-and-speed criterion ITT1. A fifth group contains ITT6, ITT7, and ITT8. It is natural to find the tests ITT7 and ITT8 in the same factor as they measure the pilots' capability in handling some of the more modern weapons. ITT6 and ITT7 both test the

[1] Due to secrecy of the material, all descriptions of tests, etc. is fictitious in this report. Factor loadings, eigenvalues, and variances are correct.

Vol. 6
1978

TABLE 2. ROTATED FACTOR MATRIX WITH TEN CRITERIA, FIVE FACTORS, AND 36 OBSERVATIONS

Criteria	Factor Number†				
	1	2	3	4	5
ITT1	0.04	−0.83**	0.08	0.38	−0.08
ITT2	0.29	−0.73*	0.14	0.37	−0.18
ITT3	0.23	−0.18	0.09	0.88**	−0.09
ITT4	0.39	−0.70*	−0.21	0.01	−0.36
ITT5	0.26	−0.27	0.06	0.11	−0.89**
ITT6	0.81**	−0.21	0.09	0.07	−0.37
ITT7	0.84**	−0.14	0.17	0.19	−0.30
ITT8	0.87**	−0.28	−0.15	0.16	0.08
ITT9	0.30	−0.73*	0.31	−0.26	−0.15
ITT10	0.02	−0.12	0.96**	0.10	−0.04
Eigenvalues	4.78	1.29	0.98	0.94	0.65
Cumulative proportion of explained variance	48	61	71	80	86

† Two asterisks denote a high factor loading and one asterisk a medium high loading.

TABLE 3. GROUPING OF THE CRITERIA AND LABELLING OF THE FACTORS

Factor	Criteria	Method of interception	Visibility	Altitude	Speed	Use of counter measures	Evasiive action	Modern weapons
4	ITT3	CCI	IMC	Medium	Low	No	No	No
2	ITT2	CCI	IMC	High	Medium	No	No	No
	ITT1	LPI	VMC	High	High	No	No	No
	ITT9	LPI	IMC	High	Medium	No	No	No
	ITT4	LPI	IMC	Medium	Medium	No	No	No
1	ITT6	CCI	VMC	Low	Medium	No	Yes	No
	ITT7	LPI	VMC at night	Low	Medium	No	Yes	Yes
	ITT8	CCI	IMC	Medium	Medium	No	No	Yes
5	ITT5	CCI/LPI	VMC at night	Medium	Medium	Yes	No	No
3	ITT10	CCI	VMC at night	High	Medium	Yes	Yes	No

pilots' capability in intercepting an evasive enemy aircraft. Thus we can label this factor "interception of evasive aircraft by modern weapons".

6. CONCLUSIONS

The objective of this paper was to demonstrate how, in practice, the number of criteria in a problem may be reduced to a more manageable quantity. The technique of factor analysis was employed as this considered all of the original criteria and enabled the decision-maker to control the amount of lost information. By judicious manipulation, the analysis produced interpretable factors. The officers concerned were then able to estimate weights rapidly, accurately and *directly* [18], although an indirect approach could have been employed.

REFERENCES

1. FISHBURN PC (1964) *Decision and Value Theory*. Wiley, New York.
2. GUTTMAN L (1954) Some necessary conditions for common factor analysis. *Psychometrica* **19**, 458–479.
3. HARMAN HH (1970) *Modern Factor Analysis* (2nd Rev. Edn). University of Chicago Press, Chicago.
4. KEENEY RL & RAIFFA H (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, New York.
5. KEENEY RL & NAIR K (1976) Evaluating potential nuclear power plant sites in the Pacific Northwest using decision analysis. International Institute for Applied Systems Analysis, PP-76-1, Laxenberg.
6. KORT F (1968) A nonlinear model for the analysis of judicial decisions. *Am. Polit. Sci. Rev.* **62**, 546–555.
7. MILLER GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**(2), 81–97.
8. TELL B (1978) An approach to solving multi-person multiple-criteria decision-making problems in *Multiple-criteria Problem Solving—Theory, Methodology, and Practice in Buffalo* 1977 (Ed. ZIONTS S), Springer Verlag, Berlin.

9. Turban E & Metersky ML (1971) Utility theory applied to multi-variable system effectiveness evaluation. *Mgmt Sci.* **17**(12), B817–B828.
10. Yntema DB & Torgerson WS (1961) Man–computer cooperation in decision requiring common sense. *IRE Trans. on Human Factors in Electronics.* **HFE-2**, 20–6.

Bertill Tell
(*March* 1978)

Assistant Professor in Management Science
Stockholm School of Economics
BOX 6501
S-113 83 *Stockholm*
Sweden

Vol. 6
1978

OMEGA, The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 455–457
© Pergamon Press Ltd. 1978. Printed in Great Britain
0305-0483/78/1001-0455$02.00/0

A Bayesian Forecasting Model for House Account Sales

THE COMPANY where this sales forecasting model has been implemented is one of the main UK producers of toiletries, forming part of a major multinational organisation. Seventy-five per cent of all company sales results from the efforts of some 150 representatives calling on about 30,000 outlets throughout the United Kingdom. The remaining 25% is taken by three large National Multiple chains, known as House Accounts. Each of these Multiples purchase their toiletries centrally through the direct negotiations of their chief buyer and the House Account Manager.

Thus the sales forecasts of the House Account Manager will be based not only upon the general market conditions but also from his personal knowledge of the special preferences and business circumstances of the particular toiletry buyer. The highly personalised nature of this buying behaviour renders much of conventional forecasting theory such as time-series and regression analysis inappropriate. Yet, the House Account Manager requires a more systematic procedure for evaluating all of the factors known to influence the sales than by just setting on a figure which "feels about right". The Sales Forecasting Model presently developed is therefore an attempt to provide an approach, soundly based upon principles of Decision Theory, which represents an intuitively acceptable synthesis of the largely subjectively assessed factors which influence house account sales. The next section describes briefly the structure of the model and this is followed by a discussion on its performance six months after implementation and some plans for future refinements.

THE SALES FORECASTING MODEL: STRUCTURE

Essentially, the model is a Composite Bayesian Predictor, based upon a synthesis of subjective probability estimates of sales and explicitly incorporating the influence of eight important factors, viz:

1. Advertising level
2. Success of previous similar deal
3. Present stock level of line
4. "Type" of deal (banded pack, "3P OFF"...)
5. Trade stim (degree of "investment")
6. Current competitive activity
7. Brand share (increasing or decreasing)
8. Brand leadership

In the first instance, it was decided to quantify these factors only insofar as they would have a relatively High, Medium or Low influence on sales. For example, consider the promotion of Brand *X* toothpaste. The House Account Manager ticks the appropriate cells in a matrix, shown in Fig. 1. The interpretation of this matrix can perhaps best be illustrated by means of an example.

Factor	*Indications*
1.	The manager already knows, or can confirm from the appropriate brand manager, the level of advertising accompanying his promotion. In the example shown in Fig. 1, this was considered relatively low.
2 & 3.	Feedback from the buyers and merchandisers will give the manager information upon how successful (from the House Account's point of view) the previous Brand *X* promotion was and also their present *X* stock position.
4.	If the buyer is known to have a particular preference for a "banded pack", then an *X* banded pack would be regarded as a HIGH type of deal. A banded pack with a genuine and substantial "FREE" content is usually very attractive to a House Account buyer. A "bargain pack" (e.g. "only 22p") is less attractive, and a "flashed pack", e.g. "3p off", is generally looked on least favourably. House Accounts tend to be wary of "price marked packs", e.g. "only 22p" since there is a possibility that a competitor could sell "regular" stock, perhaps left over from a previous promotion, at 20p. This might suggest to a customer that the competitor's regular selling price is less than this particular House Account's special offer.
5.	The level of "Trade Stim" is the degree of "investment" in the promotion as perceived by the buyer.
6.	Current competitive activity reflects the manager's awareness of other promotions, etc.
7.	Data is usually available on the change of Brand Share from the normal market research sources.

	1	2	3	4	5	6	7	8
High			✓		✓			
Medium		✓		✓		✓		
Low	✓						✓	✓

FIG. 1.

Apart from quantifying, in these very simple terms, the eight factors, the only additional input from the manager is an estimate of his subjective probability distribution of the sales. The manager is asked to envisage High, Medium or Low scenarios and attach a measure of his degree of belief in the likelihood of each in terms of probabilities (see, for example, Fig. 2)

An updated set of data is retained by the procedure in order that the conditional probabilities $P(V_j|S)$, where

S = level of sales (high, medium or low) and
V_j = level of factor j (high, medium or low) for $j = 1$ to 8,

can be estimated from the number of occasions that V_j took one of the three values when the level of sales turned out to be S. Thus, there are eight matrices of the style shown in Fig. 3, which is a hypothetical example from the "Type of Deal" factor.

With Fig. 2 effectively articulating the manager's prior subjective probability distribution on S, the application of Bayes theorem (c.f. [5], [6]) easily generates a predictor posterior to V_j, i.e.

$$P(S|V_j) \Theta P(V_j|S)\, P(S)$$

With these eight distinct Bayesian predictors of S, the theory of linear combinations [2] is then utilised to achieve the composite Bayesian predictor

$$P\left(S\left|\sum_{j=1}^{8} V_j\right.\right) = \sum_{j=1}^{8} P(S|V_j)\, P(M_j)$$

where $P(M_j)$ represents the probability that the predictor based upon factor j will outperform the other seven predictors in the sense of being more accurate. The $P(M_j)$ used are strictly speaking the posterior means of the subjective Dirichlet distribution defined on the set $\{M_j\}$. In the procedure as it has so far developed, $P(M_j)$ are evaluated on a frequency basis from the observed data. Thus $P(M_j) = X_j/n$, where X_j = the number of times the mean value of predictor j has been seen to outperform all others over the n representative instances observed. Hence, these "outperformance probabilities" and the "likelihood matrices" (e.g. Fig. 3) are constantly updated in the computer. The manager just provides the data in Figs 1 and 2 in order for the synthesised Bayesian predictor to be computed.

It should be emphasised that this development of a Sales Forecasting Model represents more of an approach to formalising an algorithm which can incorporate a large number of factors, subjective and behavioural as well as quantitative, to produce a validated and intuitively acceptable synthesised predictor, rather than a forecasting model in the more usual sense of representing a particular mathematical equation to describe the sales generating process. At present it exists in only the very simplest version of a skeleton which can accommodate greater refinements without changing the character of the approach. Only the very simplest Bayesian predictors are presently in use. Clearly it is envisaged that these can be improved individually and also extended from the present number of eight. A considerable part of current theory in marketing, advertising and time-series analysis, could, for example, be formalised and incorporated into this model without altering the basic approach of developing a set of individual Bayesian predictors and synthesising this set, again in a Bayesian manner, to provide a composite predictor.

	Sales estimate (gross)	Probability of estimate
High	1200	0.2
Medium	1000	0.4
Low	700	0.4

FIG. 2.

Level of Sales forecast	'Type' of deal HIGH	MEDIUM	LOW
High	0.7	0.2	0.1
Medium	0.25	0.5	0.25
Low	0.2	0.35	0.45

FIG 3.

Bayesian methods are necessitated in order to incorporate the high degree of subjective information associated with the extraordinary personalised nature of House Account buying and also for the method to be valid within Decision Theory. The decomposition into the separate predictors and the subsequent synthesis becomes necessary because of the estimation problems which would arise if all possible variables were put into one equation.

IMPLEMENTATION AND FUTURE DEVELOPMENTS

The model was originally developed as described during the early part of 1975. After an extensive attempt to investigate how it would have performed on the historical data, the results were sufficiently encouraging for the model to be introduced "live" alongside the existing sales estimation procedure towards the end of 1975. The "live" performance over 30 promotions on one House Account has so far been seen to reduce the mean deviation of the forecast value from the actual from 62% to 48%, i.e. it appears to be 23% more accurate. Incidentally, the apparently large forecast errors are due to the relatively large lead time involved in these particular forecasts.

Two main refinements are currently envisaged:

(i) The use of the three point estimates in order to encode the manager's subjective probability distribution on the sales has not been entirely satisfactory. There has been an apparent tendency for the managers to underestimate the probabilities of the extremes of the range. This is an example of Anchoring Bias, discussed for example by Tversky [7], and Bunn [3]. We plan to replace the point estimates by means of the fractile assessment procedure using tertiles. Barclay and Peterson's experiments [1] suggest that the tertile method might be the best of the simplest procedures for avoiding excessive anchoring bias. This would involve the manager stating two points on the range of possible sales values which split it up into 3 equally likely intervals. Thus, instead of H, M and L point estimates we would then have Higher, Medial and Lower ranges and, after modifying the values in the likelihood matrices, this would involve very little change in the operation of the model.

Vol. 6
1978

(ii) The use of the Dirichlet distribution on (M_j) is wasteful of information. Only the relative probability of the model which performs best is revised. No account is taken of the relative performance of all the models. More efficient procedures for revising these probabilities are now available (c.f. [4]) and will be incorporated. They only represent a refinement of the computer software and not the visible operation of the model.

In all, the development of this Sales Forecasting Model has been well-received by the managers concerned. The approach has the intuitive appeal of a logical formalisation of what previously was a purely subjective process and its performance has so far evidently been a significant improvement in forecasting a difficult set of variables.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the considerable help of Dr RGM Bolton and the encouragement of Peter Murray.

REFERENCES

1. BARCLAY S & PETERSON CR (1973) Two methods for assessing probability distributions. Technical Report 73-1, Decisions and Designs Inc.
2. BUNN DW (1975) A Bayesian approach to the linear combination of forecasts. *Opl. Res. Q.* **26**(2i), 325–329.
3. BUNN DW (1975) Anchoring bias in the assessment of subjective probability. *Opl. Res. Q.* **26**(2ii), 449–454.
4. BUNN DW (1977) A comparative evaluation of the outperformance and minimum variance procedures for the linear synthesis of forecasts. *Opl. Res. Q.* **28**(3ii), 653–662.
5. LINDLEY DV (1970) *Making Decisions*. Wiley-Interscience, London.
6. THOMAS H (1972) *Decision Theory and the Manager*. Pitman, London.
7. TVESKY (1974) Assessing uncertainty. *JL R.Statist. Soc. B.* **36**(2), 148–159.

Dr Derek W Bunn

(April 1978)

University of Southern California
School of Business Administration
Department of Quantitative Business Analysis
University Park
Los Angeles
CA 90007
USA

OMEGA, The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 458–459
© Pergamon Press Ltd. 1978. Printed in Great Britain

0305-0483/78/1001-0458$02.00/0

Computers and Privacy

"... Computers *are* there to elevate man's intellect and increase his control over his environment".

H. SACKMAN, *Computers, System Science and Evolving Society*, Wiley (1967)

THE REPLACEMENT of the word "are" by the expression "should be" would give a more accurate assessment of the present state of computing. Enid Mumford [1] has highlighted the fact that the present philosophy associated with the design of computer systems is too technically oriented. There is little evidence that the introduction of computers has actually increased man's control over his environment. Certainly there has been an information or data explosion, but too much data is almost as bad as too little.

What is important and cannot be emphasised too strongly is that man must control the use to which computers are being put. The present almost anarchic state in which man freely blames his computers for almost everything is certainly more attractive than the kind of society portrayed so graphically by George Orwell [2] in his novel "1984". Nevertheless it is far from satisfactory.

Indeed it is perhaps an apposite moment to try to answer the question "are we any nearer a '1984' type of society as a direct result of using computers?". Many observers have detected trends towards the kind of society envisaged by Orwell. A vast increase in the powers of the national state and supranational bodies such as the EEC have little by little resulted in a steady decline in the importance of the individual. True freedom for any individual has, of course, never existed. Indeed such an anarchistic society holds little attraction to rational man.

There are in addition other threats to freedom, more subtle perhaps, but threats nevertheless, and intermingled with many of these, is that little understood child of the 20th Century, the electronic computer.

First we must set the record straight. The computer is not "a deep and mysterious machine" [3], nor is it a "bundle of wires and cogs" [4]. It may contain a wire or two and perhaps even a cog but it is however a machine which is capable of reproducing man's instructions at speeds which appear to belong to the realm of science fiction. It can serve as a calculating machine, the latest in a long line of aids to 'arithmetic man', it can serve as a library of information, it can serve as a payroll clerk. In short it can act as an information processor.

When computers are used within large industrial corporations or large bureaucracies such as Departments of State, reference is often made to the computer as being part of a Management Information System (MIS). The term MIS came of age about eight to ten years ago just at the time when the use of computers for fairly simple administrative tasks began to become commonplace. Dreamy computer men, bored with payrolls and accounts receivable, began to design vast networks of central processors, terminals and magnetic storage devices. Instead of rushing to a filing cabinet, the secretary would have to learn to interrogate a computer-based file. Managing directors would be able to speculate as to the company's likely cash position in the future in the event of crisis A or following decision B. It must be emphasised that none of this was completely new ground. Managing Directors are supposed to speculate in this way, but being human, they rarely find the time or the inclination to grapple with the necessary arithmetic. Such tasks are traditionally delegated to staff personnel. The arrival of the terminal on the manager's desk would change all this.

Other people went further than the dreamy computer professionals. The still fledgling discipline of operational research produced the horrifying concept of centralised computerised control of organisations. What was good for a power station ought to be good enough for the office. Too much talk of the impending computer revolution led not surprisingly to a strong reaction from a number of organisations and individuals who were ostensibly interested in the freedom of man. The computer or rather its bank of magnetic memories became another bogey man, like the much maligned man in blue. Society obviously needs a bogeyman. Napoleon has gone, so has Hitler and somebody or something was needed to threaten the freedom of the little man. Computers were already notorious for such frivolities as demands for £0.00. Incorrect bank statements were around in numbers which old fashioned non-computerised managers would never have thought possible. The news that the Inland Revenue in the UK were thinking of an extensive investment in computers must have brought smiles if not "three cheers" from the hardened tax evader. However, the prospect of the Inland Revenue computer talking to the education computer or the social security computer inspired men to believe that "1984" was around the corner. A closer analogy might be the Tower of Babel.

In practice there are three principle dangers to privacy:

1. Inaccurate, incomplete or irrelevant information.

2. The possibility of access to informattion by people who should not or need not have it.

3. The use of information in a context of for a purpose other than that for which it was collected.

They are all very real dangers and it is to be hoped that Governments and Management will treat them seriously but not hysterically as unfettered "privacy" can also be dangerous. To conclude, another quotation, which surely will bring heart to all computer men: "Science of Calculation... which must ultimately govern the whole of the application of science to the Arts of Life". (Charles Babbage).

REFERENCES

1. MUMFORD E (1973) Designing systems for job satisfaction. *Omega* **1**(3), 493–498.

2. ORWELL G (1949) 1984. Secker and Warburg, London.
3. LORD SNOW (1966) *Hansard* (March 2). House of Lords.
4. WORSTHORNE P (1965) *Weekend Telegraph* (December 4). London.

Dr JO Jenkins

Department of Management Science
Imperial College
Exhibition Road
London SW7 2BX
UK

(*May* 1978)

l. 6
78

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 460–461
© Pergamon Press Ltd. 1978. Printed in Great Britain

0305-0483/78/1001-0460$02.00/0

On the Design and Control of Crisis Games

INTRODUCTION

CONFLICT occurs when two or more systems are perceived as pursuing aims or policies some of which are mutually incompatible; crisis occurs when a system is unable to cope with its intended sphere of activity [1]. In a decision making context, a crisis involves a threat to important values; surprise, so that there are no suitable plans for coping with the crisis; and a short time in which a decision must be taken [2]. The behaviour of individual decision makers in conflicts and crises is of some interest, and experimental gaming provides a useful approach for its investigation. This note introduces the Superior Commander methodology for the design and control of crisis games.

REQUIREMENTS FOR A CRISIS GAME

The elements of threat, surprise and short time to act which characterise crises are stress-producing stimuli, or stressors [3]. The effects of these stressors on a Player in a game may differ according to the role he thinks he is playing. In his role as Decision Maker in the game, the Player's behaviour may change, and this is of interest to us. (For examples of the kinds of changes we might observe, see [2, Appendix].) The Player also has the role of Experimental Subject, and the effects of stress on him in this role may also be important. For example, one indication of stress is objective averseness to the stressor, manifested as escape or avoidance behaviour. One way for the Player to avoid the stressor is to change his behaviour as Experimental Subject, to regard the game as less important, and to become less committed to his role as Decision Maker. Effects of this kind may be similar to those we want to investigate: they are not wanted, and a useful gaming methodology will attempt to minimise them. The Player must believe that he is an active *Decision Maker* with an important task to perform in the game.

The scientific testing of specific hypotheses requires that the game be *replicable*, so that different subjects can be placed in the same game situations. Then it should be possible to examine quantitative hypotheses under tight experimental control.

Finally, the game must be sufficiently complex and detailed for the player to regard it as reasonably realistic. Then we could be more confident that his actions and decisions in the game were not peculiar to the game-world, but corresponded to his behaviour when confronted by real decisions and crises. These requirements may be summarised:

R1: The Player in a crisis game must believe that he is an active Decision Maker.

R2: A crisis game must contain sufficient complexity and detail to be accepted by the Player as realistic.

R3: The experimenter must be able to control a crisis game so that he can replicate it exactly with different subjects.

We note that these requirements appear to be conflicting. To satisfy R3, a simple game is indicated, but a simple game would not provide the complexity and detail needed for R1 and R2. The Superior Commander methodology is a scheme which reconciles these aspects.

THE ORGANISATIONAL CONTROL GAME

We have designed a game, the Organisational Control Game, as an implementation and example of the Superior Commander system. Experimental subjects are told that the game is designed to investigate decision making in a hierarchical system.

A conflict between Blue and Red forces is depicted in a form of a wargame. Blue forces are controlled at three levels (Fig. 1). A Corps Commander, with a very general map, has overall control of three Divisions, commanded by different players. Each Divisional Commander (Div Comd) has a more detailed map of his area on which battalion-sized units appear; three Task Forces (TFs) are controlled by each Div Comd. The TFs each have individual Commanders, who engage in combat with the opposing Red forces at the level of company-sized units; combat is resolved according to the game rules, as in a conventional wargame (see, for example, [4]).

Normally the players are isolated from one another, in separate rooms. Communication takes place in writing, and messages are transmitted via Game Control. For example, the structure of the game, *from the perspective of a Div Comd*, is shown in Fig. 2.

THE SUPERIOR COMMANDER SYSTEM

The actual game, *as the experimenter sees it*, is not like this (Fig. 3). All players are Div Comds; there are no Corps or TF Comds, and players communicate only with Game Control.

The entire game is predetermined by Game Control, and the actions of the players have no effect on its outcome. The scenarios are designed so that the range of options for the Blue forces is small; any orders the Div Comd issues to his TF Comd, if they are tactically reasonable, are likely to be 'obeyed', because the TF

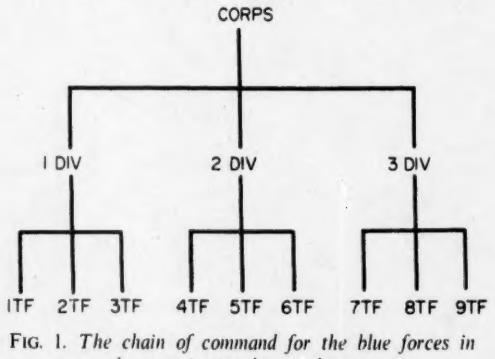


FIG. 1. *The chain of command for the blue forces in the organisational control game.*

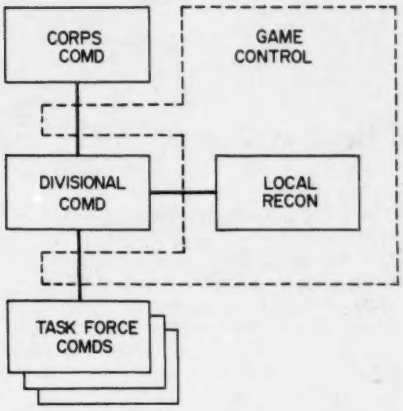


FIG. 2. *The structure of the game, as a divisional commander sees it.*

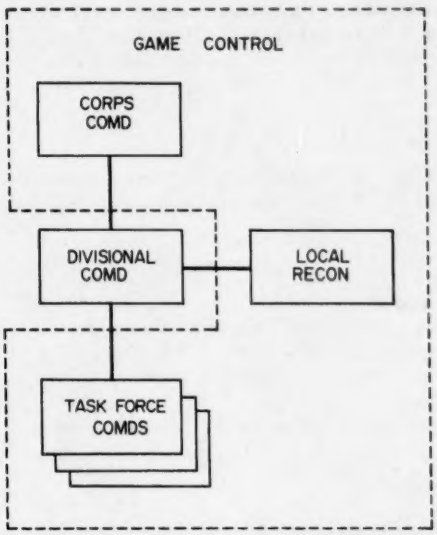


FIG. 3. *The structure of the game, as the experimenter sees it.*

Comds have been pre-programmed to react sensibly. Any failures to obey the Div Comd's orders can be explained in terms of combat at the lower level, details of which are not apparent on the Div Comd's map.

The Div Comd is responsible for the allocation of air strikes to the TFs. This gives him a task to perform, a task which appears to him to be tactically important, but which in fact has no bearing whatever on the outcome of the game. This dummy task is important in maintaining the player's involvement as a Decision Maker in the game, and it helps to maintain the deception.

Much complexity and detail can be included in the game, and the hierarchical command structure is easy for the player to accept in this context. Thus our three requirements for a crisis game are met: the player has an active role which he sees as significant in the game (R1); the game is complex and detailed, and reasonably realistic (R2); and the game, because it is predetermined, can be replicated exactly with different players (R3). Initial testing of the game has been most encouraging.

SUMMARY AND CONCLUSIONS

The investigation of the behaviour of individual decision makers under conditions of crisis and stress imposes special requirements on an experimental game. The Superior Commander is a methodology which we claim satisfies these requirements. It provides an environment in which realistic, detailed games be played and replicated, under close experimental control. It is designed so that the player thinks he is an active decision maker in the game, thus to some extent minimising the undesirable aspects of stress associated with his two separate roles, as Decision Maker in the game and as Experimental Subject.

The Superior Commander methodology is not restricted to the investigation of behaviour in crises, however; it is a general methodology for research games. It can be used for a range of purposes for which realism and good experimental control are both required.

ACKNOWLEDGEMENTS

The advice and encouragement of Mr KC Bowen is gratefully acknowledged. This research is funded by the Ministry of Defence under Contract Number ER3/9/4/2088/012 to Professor MRC McDowell.

REFERENCES

1. HARRIS JI (1974) Some concepts of a conflict game. Defence Operational Analysis Establishment, *Memorandum* **7420**, Ministry of Defence, West Byfleet, UK.
2. HERMANN CF (Ed.) (1972) *International Crises: Insights from Behavioural Research*. Free Press, New York; Collier-Macmillan, London.
3. MILBURN FW (1972) The management of crisis. In *International Crises* (ED. HERMANN CF). Free Press, New York, Chapter 11.
4. PALMER N (1977) *The Comprehensive Guide to Board Wargaming*. Arthur Barker, London.

Dr Dale F Cooper

(*May* 1978)

Operational Research Group
Mathematics Department
Royal Holloway College
Egham
Surrey TW20 0EX
UK

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 5, pp. 462–464
© Pergamon Press Ltd. 1978. Printed in Great Britain

0305-0483/78/1001-0462$02.00/0

Capital Budgeting: a Sequential Decision Model

THE BASIC capital budgeting model is:

$$\text{Max} \sum_{i \in I} r_i x_i \quad \text{subject to}$$

$$\times \sum_{i \in I} d_{it} x_i \leq b_t, \; t = 1, \ldots T \quad \text{and} \quad \forall x_i = [0, 1]. \quad (1)$$

This model describes a firm faced with selecting among I independent and competing projects. For the ith project, its value to the firm is given by r_i and its demand for capital over the T budget periods is given by $(d_{i,t}, t = 1, \ldots T)$. The presence of a fixed capital budget, $(b_t, t = 1, \ldots T)$, forces the firm to discriminate where the ith project may be accepted ($x_i = 1$) or not ($x_i = 0$). This formation is referred to as the Lorrie Savage Model [5]; the reader is referred to a state-of-art survey, (e.g. [1], [6] for more comprehensive information.

The model in (1) and subsequent extensions assume the decision maker has a complete demand function at the time of decision. That is, all of the projects to compete for the budget are present at a fixed point in time. However, experimental evidence on the evaluation of capital projects [2] suggests that this process will often be applied sequentially due to the fact that demand for project funding is spaced throughout a budget period and because management will address components of demand as they arise. This paper departs from previous mathematical programming models on capital budgeting by presenting a formulation which recognizes the need for rationing capital while dealing sequentially with project evaluation over the budget period. The formulation is a stochastic programming model which considers the competition for funds not only from existing project proposals but also from proposals yet to arrive during the budget period. The model allows for recourse with current project proposals in that if a specific project is not accepted now then it may be accepted later in the budget period.

DECISION MAKING ENVIRONMENT

Consider a single budget period and a budget that has been set for capital projects over that period. Projects requesting funds from the budget will arise sequentially over the period and a project evaluation process will be enacted a number of times within the period. During each review process, each project will either be accepted or deferred. The timing of reviews may be periodic or it may be random, being triggered whenever a new project requesting funds is introduced or more information is obtained relating to the attractiveness of a project previously introduced but yet to be accepted. Each review process will be shaped by the following elements: (i) the time remaining in the budget period, (ii) the portion of the original budget which has yet to be committed to projects, (iii) a set of projects requesting funding and (iv) a forecast on the quality and capital requirements of projects likely to still arrive within the current budget period.

In determining whether a specific project should be undertaken the decision maker must assess the project's competitive position with respect to the other projects under review, (i.e. standard capital budgeting model) and also assess the project's position with respect to other projects which could possibly arrive before the budget period is over. Several opportunity costs must be considered. Accepting a good project now may prevent the firm from accepting an even better project later in the budget period. Deferring action on a specific project will result in opportunity costs if the project is found to be acceptable at a subsequent review in the budget period. This cost may be moderate if the result is a simple time delay in receiving the project's value or the cost could be major if because of the time delay the project becomes infeasible. Accordingly, the capital budgeting problem being addressed is not a static process of accepting and rejecting projects. The process is dynamic and management's goal is to get the best possible projects into operation as early as possible.

MODEL FORMULATION

Consider an arbitrary review process within the budget period. The following assumptions are now made:

Other project arrivals and reviews yet to occur within the period are assumed to occur at a single random point in time. This enables current review to be treated as a two stage decision process and as will become evident, yields a computationally tractable model.

For any project under review there is a chance of losing the project if it is deferred; this is modeled as a simple Bernoulli random variable.

For any project yet to arrive, and which is deferred to the next budget period, there is the chance of losing the project. It is assumed that future projects, (i.e. unknown projects) can only be forecasted on an aggregate demand level; consequently any loss factor must be introduced on this basis. A single deterministic parameter will be used to specify losses for any fraction of future project demand which is deferred to the next budget period.

So as not to distort the economic impact of deferring projects beyond the current budget period an end effects assumption is necessary. This formulation assumes that all unanswered demand in the current budget period will be met at the beginning of the next budget period. Accordingly, a single time value discount parameter may be used to represent the impact of this delay. Also, it is assumed that any demand which is lost, while being deferred, can be replenished in the next period but the quality (i.e. value to the firm) of the replacement demand is given by the mean quality level over all project demand.

The firm's objective is to maximize the value of the remaining budget and any carryover value of unanswered demand going into the next period.

Model Parameters

τ = time remaining in the budget period.
RB = uncommitted portion of original budget.
I = number of specific projects presently under review.
v_i = value of the ith project to the firm.
d_i = demand for capital from the ith project.
β_i = probability of losing the ith project if it is deferred.
s_i = a 0/1 Bernoulli random variable with probabilities β_i and $(1-\beta_i)$, respectively; s_i models the availability of the i-th project during τ given the project is deferred under the current review.
$(q_k, k = 1,\ldots, K)$ = K discrete quality levels characterize the range of projects yet to arrive during τ where project quality is defined as the value to the firm per unit of capital required.
$(h_k, k = 1,\ldots, K)$ = the aggregate capital demands from projects yet to arrive during τ are specified for each discrete quality level. Such demands must be forecasted and the model assumes that each is a discrete random variable.
γ_1 = discount applied to the value of any project whose acceptance occurs at some future time in the budget period.
γ_2 = discount applied to the value of any project whose acceptance is delayed until the next budget period.
α = fraction of this period's future demand to be lost if not satisfied until the next budget period.
$\bar{q}$ = average quality level from all project demand over an entire budget period.
$g(s_i)$ = value of the ith project (before time discounting) when deferred to the next budget period; $g(s_i) \equiv (1-s_i)\,\bar{q}d_i + s_i v_i$.
$f(q_k)$ = per unit value of demand from the kth quality level (before time discounting) when demand is deferred to the next budget period; $f(q_k) \equiv (1-\alpha)\,q_k + \alpha\,\bar{q}$.

Decision Variables

$(x_i, i = 1,\ldots, I)$ = accept/defer decisions on projects under review; x_i is 1 or 0 depending on whether the ith project is accepted now or deferred to subsequent reviews during τ.
$(x'_i, i = 1,\ldots, I)$ = accept/defer decisions on current projects deferred to subsequent reviews during τ; if the ith project is presently deferred, (i.e. $x_i = 0$) then x'_i is 1 or 0 depending on whether the ith project is accepted in subsequent reviews or deferred to the next budget period.
$(y_k, k = 1,\ldots, K)$ = capital to be allocated for projects yet to arrive during τ.

Stochastic Programming Model

$$\text{Max} \sum_{i\in I} v_i x_i + \gamma_1\left[\sum_{i\in I} v_i x'_i + \sum_{k\in K} q_k v_k\right] + \gamma_2\left[\sum_{i\in I} g(s_i)(1 - x_i - x'_i) + \sum_{k\in K} f(q_k) \times (h_k - y_k)\right] \quad (2)$$

subject to:

$$\sum_{i\in I} d_i(x_i + x'_i) + \sum_{k\in K} y_k \le RB \quad (3)$$

$$x_i + x'_i \le 1: \quad i = 1,\ldots, I \quad (4)$$

$$x'_i \le s_i: \quad i = 1,\ldots, I \quad (5)$$

$$y_k \le h_k: \quad k = 1,\ldots, K \quad (6)$$

$$x_i, x'_i \text{ are } [0, 1]\forall i \quad (7)$$

Program (2–7), reviews the current projects with the aim of maximizing the firm's utilization of the remaining budget. The objective function has three distinct components: value derived from accepting projects now, value from accepting projects later in the period and value from accepting any of this period's projects at the beginning of the next period. Constraint (3) is for capital budgeting; (4) restricts a project from being taken both now and later in the period; (5) restricts a project from being accepted later on the chance it is lost while being deferred; (6) controls the level of possible investment into projects yet to arrive and (7) are simply binary constraints which recognize the indivisibility of the current projects. The above program is stochastic due to the discrete random variables, s (future availability of current projects) and h (future project demand). The primary decisions variables are x and the recourse available to the firm is represented by the secondary decision variables x' and y and by the end effects assumption.

MODEL IMPLEMENTATION

The discrete random variables, s and h may be represented by a finite set of outcomes, $[s_{i,n}, i = 1,\ldots, I), (h_{k,n}, k = 1,\ldots, K), n = 1,\ldots, N]$ where the probability of the n-th outcome is denoted by p_n. Given that it is acceptable to maximize the expected value of (2), the stochastic program may be replaced by the deterministic equivalent.

$$\text{Max} \sum_{n\in N} p_n\left[\left[\sum_{i\in I} \{[v_i - \gamma_2 g(s_{i,n})]x_i + [\gamma_1 v_i - \gamma_2 g(s_{i,n})]x'_{i,n}\} + \sum_{k\in K} [g_k - \gamma_2 f(q_k)]y_{k,n}\right]\right] \quad (8)$$

subject to:

$$\sum_{i\in I} d_i(x_i + x'_{i,n}) + \sum_{k\in K} y_{k,n} \le RB \quad n = 1,\ldots, N \quad (9)$$

$$x_i + x'_{i,n} \le 1 \quad i = 1,\ldots, I: \quad n = 1,\ldots, N \quad (10)$$

$$x'_{i,n} \le s_{i,n} \quad i = 1,\ldots, I: \quad n = 1,\ldots, N \quad (11)$$

$$y_{k,n} \le h_{k,n} \quad k = 1,\ldots, K: \quad n = 1,\ldots, N \quad (12)$$

$$x_i, x'_{i,n} \text{ are } [0, 1]\forall i, n \quad (13)$$

Realistically, N will be quite large if the probability functions on s and h are adequately modeled. This implies a large scale program which is further complicated with a large number of binary decision variables, $(x'_{i,n}, i = 1,\ldots, I; n = 1,\ldots, N)$. For example, assume there are five projects under review, (i.e. $I = 5$). If only three project quality levels, (i.e. $K = 3$) are used in characterizing future demand and if each of these demand levels is modeled by a three point probability function then, assuming the elements of (s, h) have probabilitic independence, $N = 864$ $(N = 2^I 3^K)$ and the number of binary variables equals 4325. Any attempt to solve the program directly is likely to result in an impractical decision making tool. However, the program does have a special structure and solutions may be more easily obtained. By projecting on the primary decisions x, the program decomposes into N trivial sized mixed integer subprograms (see [3] for details) of the form:

$$\text{Max} \sum_{i \in I} [\gamma_1 r_i - \gamma_2 g(s_{i,n})] x'_{i,n} + \sum_{k \in K} [\gamma_1 q_k - \gamma_2 f(q_k)] y_{k,n} \quad (14)$$

subject to:

$$\sum_{i \in I} d_i x'_{i,n} + \sum_{k \in K} y_{k,n} \leq RB - \sum_{i \in I} d_i x_i \quad (15)$$

$$x'_{i,n} \leq \text{Min.}\,[(1 - x_i), s_{i,n}] \quad i = 1,\ldots, I \quad (16)$$

$$y_{k,n} \leq h_{k,n} \quad k = 1,\ldots, K \quad (17)$$

$$x'_{i,n} \text{ is } [0, 1] \forall i \quad (18)$$

The computational effort for solving families of these subprograms within an iterative search on x should be quite manageable.

Any assessment of program implementation also requires consideration of the data input requirements. The program formulation requires several inputs which are non-standard to previous capital budgeting models: (i) the loss probabilities $(\beta_i, i = 1,\ldots, I)$ for s and the aggregate loss factor α and (ii) the probability functions on future project demand $(h_k, k = 1,\ldots, K)$ and $\bar{q}$. In assigning a value to β_i (the probability of losing the ith project if acceptance is deferred) the project originator will have to employ subjective probability estimation techniques (see, for example, [4]).

For any given review, the model assumes there will be only one additional review in the period, (i.e. a two stage model). In practice, there will probably be several more reviews and the probability of losing a project is sure to depend on how long the project is delayed. Accordingly, averaging techniques with respect to the length of a project delay will be necessary in forming a single loss probability for each project. The aggregate loss factor α can be established by averaging historical results on all projects lost while under delay.

Securing probability functions on $(h_k, k = 1,\ldots, K)$ is a forecasting problem and the application of time series analysis appears ideal for providing this data. The mean project quality level, $\bar{q}$ can be set by taking an expectation over $(h_k, k = 1,\ldots, K)$ which is properly weighted by the K discrete quality levels.

In summary, the basic capital budgeting problem (1) has been extended to model capital rationing in a sequential environment where competition from future projects is systematically considered along with the competition between existing projects. The formulated model is a stochastic program which appears quite reasonable for implementation.

REFERENCES

1. Bernhard RH (1971) Mathematical programming models for capital budgeting—a survey, generalization, and critique. *J. Fin. and Quantit. Analysis* **VI**(1).
2. Conrath DW (1973) From statistical decision theory to practice: some problems with the transition. *Mgmt Sci.* **19**(8).
3. Geoffrion A (1970) Elements of large-scale mathematical programming. *Mgmt Sci.* **16**(11), 652–691.
4. Hogarth RM (1975) Cognitive processes and the assessment of subjective probability distributions. *J. of the Am. Statist. Ass.* **70**, pp 271–294.
5. Lorie JH & Savage LJ (1955) Three problems in rationing capital. *J. Bus.* **XXVIII**(4), 229–239.
6. Mao JCT (1970) Survey of capital budgeting: theory and practice. *J. Fin.* **25**, pp 349–360.

Frank Noonan

(*June* 1978)

Associate Professor
Department of Management
Worcester Polytechnic Institute
Worcester
Massachusetts 01609
USA

OMEGA. The Int. Jl of Mgmt Sci., Vol. 6, No. 5, pp. 465–468, 1978. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

Bergeron M (1975) *French*: *Analyse comparative du financement des coopératives et companies dans le secteur des meuneries au Québec.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Besnehard M (1975) *French*: *Etude de l'influence du type psychologique et du type de système sur l'attitude du décideur vis-à-vis d'un système d'aide à la décision stratégique.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Boyd L (1975–76) Planning Models for General Insurance. MPhil Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Chabot-Corribeau D (1975) *French*: *Les sources d'information impersonnelles par rapport au risque objectif et à certaines caractéristiques individuelles telles que l'éducation et le revenu.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Gamonnet F (1975) *French*: *Une approche conceptuelle du comportement d'achat*: "*Les effets de la perception du produit sur le comportement des acheteurs en milieu industriel*". MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Genot J-N (1976) *French*: *Approche stochastique de la fréquentation des centres commerciaux.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Gerrard MJ (1975–76) The Quadratic Assignment Problem and Plant Layout. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Hoffman J (1971–72) *German*: *Die Besteuerung der Landwirtschaft in der EWG—Belastungsunterschiede und die Frage der Wettbewerbsverzerrungen.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Huch B (1971–72) *German*: *Das optimale Sortiment wachsender Industrieunternehmungen—unter besonderer Berücksichtigung des Markenartikels.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Jaeck H–J (1971–72) *German*: *Umrisse einer feldtheoretischen Raumkonzeption im Rahmen der Standorttheorie und der regionalen Absatzlehre für urbane Einzelhandelsagglomerationen.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Kaempf J (1971–72) *German*: *Verkehrstheoretische Uberlegungen zur Effizienzsteigerung des europäischen Linienluftverkehrssystems.* PhD Thesis, Georg–August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Klauss H (1971–72) *German*: *Zur Rehabilitation psychisch Behinderter—Eine Untersuchung unter besonderer Berücksichtigung des norddeutschen Raumes.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Korman SM (1975–76) Graph Colouring and Related Problems in Operation Research. PhD Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Kuhnemund K (1971–72) *German*: *Bikausale Deckungsbeitragsrechnung unter besonderer Berücksichtigung der Anlagendeckungsrechnung.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Nidersächsischen Staats-und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Lagrange E (1976) *French*: *Recherche de statut et consommation anticipée.* MSc Thesis. Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Laulanie F (1974) *French*: *Caractérisation du comportement boursier de deux groupes de firmes du N.Y.S.E. ayant subi des fusions de type integré et diversifié.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Lemieux G (1974) *French*: *Le financement de l'industrie de l'éducation au Canada.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Lohmann K (1971–72) *German*: *Nichtlineare Modelle zur finanziellen Leistungsprogrammplanung von Kreditinstituten.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Mandeng P (1971–72) *German*: *Die Arbeitskräftebeschaffung in den Südbezirken Kameruns während der deutschen Kolonialherrschaft (1884–1914).* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Mandron A (1975) Simulation of an Outpatient-Department Clinic. MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Mauersberger V (1971–72) *German: Rudolf Pechel un die 'Deutsche Rundschau'—Eine Studie zur konservativ-revolutionären Publizistik in der Weimarer Republik (1918–1933).* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Monier P (1976) *French: Types de financement et phases de croissance des entreprises.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Muller J-B (1971–72) *German: Bedürfnis und Gesellschaft—Bedürfnis als Grundkategorie im Liberalismus, Konservatismus und Sozialismus.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Nagachandran S (1975–76) The Economic Assignment of Men to Machines. MPhil Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

Page J-P and Prefontaine J (1974) An Empirical Analysis and Testing of Complete Portfolio Diversification within the Markowitz Efficiency Criteria. MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Petit GJM (1974) *French: La consultation et le développement de la moyenne entreprise.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Poggi G (1975) *French: Le développement organisationnel (D.O.).* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Rock R (1971–72) *German: Genesis und Analyse der Hauszeitschrift.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Scaraffioti J (1974) *French: La notion de risque: une tentative de validation.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Schluter J (1971–72) *German: Studien zum Problem der bankbetrieblichen Sicherheit.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- and Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Schmiederer R (1971–72) *German: Zur Kritik der Politischen Bildung—Ein Beitrag zur Soziologie und Didaktik des politischen Unterrichts.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Schulz H (1971–72) *German: Der Einfluss der Kapitalerhöhung aus Gesellschaftsmitteln auf die aktinkurse—Eine empirische Untersuchung für den Zehn-Jahres-Zeitraum vom 1.1.1960 zum 31.12.1969.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Seghal M (1974) *French: Le risque systématique des industries canadiennes et américaines.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Synwoldt R (1971–72) *German: Landwirtschaftliche Versicherungen als Gegenstand der Agrarpolitik.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Tennstedt F (1971–72) German: *Probleme der Berufsunfähigkeitsrenten in Deutschland.* PhD Thesis, Georg-August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Waters CDJ (1975–76) Computer-Based Facility Scheduling. MPhil Thesis, Imperial College, London. (Available at: Lyon Playfair Library, Imperial College, London, UK)

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

Blakely JL (1972) The Relative Influence of Formal and Informal Organizational Structure of Learning. SM Thesis, MIT, Cambridge, MA.

Burger RT Jr (1972) An Interactive Decision System to Support Capital Investment Decisions under Risk. SM Thesis, MIT, Cambridge, MA.

Butts REH (with **Howard EV**) (1972) Assimilation of Minority Managers. SM Thesis, MIT, Cambridge, MA.

Carter JD (1972) Long-Range Planning Model for a Consumer Food Company. SM Thesis, MIT, Cambridge, MA.

Casella RF (1972) Evaluation of the Measurement System Used in Determining the Effectiveness of Executive Development Programs. SM Thesis, MIT, Cambridge, MA.

Ciminero GL (1972) The Implementation of Alternative Investment Functions in the Brookings Econometric Model of the U.S. SM Thesis, MIT, Cambridge, MA.

Colman WT (1972) A Multistage Inventory Control Case Study in a Mass Production Factory in the United Kingdom. SM Thesis, MIT, Cambridge, MA.

Craig CE (with **Harris RC**) (1972) Productivity Concepts and Measurement—A Management Viewpoint. SM Thesis, MIT, Cambridge, MA.

D'Aversa JS (1972) Optimal Equipment Maintenance, Rebuilding and Replacement Policy. SM Thesis, MIT, Cambridge, MA.

HECHT JR (with KOCH AL) (1972) The Impact of Intercultural Sensitivity Training on the Organization and its Members. SM Thesis, MIT, Cambridge, MA.

HOMONOFF RB (with MULLINS DW) (1972) Cash Management—Applications and Extensions of the Miller–Orr Control Limit Approach. SM Thesis, MIT, Cambridge, MA.

KARP ES (1972) The Development of a Self-Selection–Orientation Workshop for the Undergraduate Studies Program. SM Thesis, MIT, Cambridge, MA.

KASTNING WH (1972) A Productivity Analysis of a Warehousing Supply System. SM Thesis, MIT, Cambridge, MA.

KATZ GM (1972) An Operational System Design for Cervical Cancer Screening in a Public Health Environment. SM Thesis, MIT, Cambridge, MA.

KIRBY MG (1972) The Lease–Purchase Phenomenon in the Capital Goods Market. SM Thesis, MIT, Cambridge, MA.

KLAPFISH MS (1972) A Simulated Forecast of the Joint Frequency Distribution of any Four-Magazine Campaign Using Readership Survey Data. SM Thesis, MIT, Cambridge, MA.

KLURFELD LF (1972) Maximizing Communications in a R and D Laboratory Through Computerized Relative Allocation of Facilities. SM Thesis, MIT, Cambridge, MA.

KOHORN Z (1972) Marketing Computer Technology to Hospitals: A Study of the Innovation–Decision Making Process. SM Thesis, MIT, Cambridge, MA.

KRIEGER AM (1972) Sampling with Missing Observations from a Bivariate Normal Distribution. SM Thesis, MIT, Cambridge, MA.

KRIVICKY GMA (1972) East–West Trade: A Case Study of the Polish and American Experience. SM Thesis, MIT, Cambridge, MA.

LAIRD MW (1972) Natural Resources and International Conflict: A Japanese Case Study. SM Thesis, MIT, Cambridge, MA.

LEJEUNE PY (1972) How Do Managers Operationalize Goals? SM Thesis, MIT, Cambridge, MA.

LENOT MR (1972) National Development Banks in Advanced Countries and their Possible Application to the United States of America. SM Thesis, MIT, Cambridge, MA.

LEONARD RJ (1972) An Empirical Examination of a New General Equilibrium Model for Warrant Pricing. SM Thesis, MIT, Cambridge, MA.

LEWENT JC (1972) The French Open-End Investment Companies (1964 to 1971): An Application of the Capital Asset Pricing Model. SM Thesis, MIT, Cambridge, MA.

LYNESS FK (1972) The Marketing of a Small Management Journal: A Case Study. SM Thesis, MIT, Cambridge, MA.

LYNN L (1972) Modelling the Intern Selection Process. SM Thesis, MIT, Cambridge, MA.

MANKOWSKI N (1972) Testing and Simulation in Portfolio Management: Framework for a Supporting Computer System. SM Thesis, MIT, Cambridge, MA.

MARLAR CH (1972) Innovations in Petroleum Marketing. SM Thesis, MIT, Cambridge, MA.

MATSON JR (1972) Future Directions in Education: A Forecast of the Impact of New Technology on Universities. SM Thesis, MIT, Cambridge, MA.

MCMAHON JT (1972) An Analysis of the Rationalization, Coordination, and Decentralization of Federal Grant Making Agencies in New England, 1969–1972. SM Thesis, MIT, Cambridge, MA.

MENEZES VJ (1972) Shadow Prices for Investment Analysis in Developing Countries. SM Thesis, MIT, Cambridge, MA.

MICHELS WC (1972) An On-Line Production Planning System. SM Thesis, MIT, Cambridge, MA.

MILLER BE (1972) A Comparison of Some Attitudes and Work Habits of Tenured and Non-Tenured Professors. SM Thesis, MIT, Cambridge, MA.

MILLER DG (1972) Pricing Policies for an In-House Computer Facility. SM Thesis, MIT, Cambridge, MA.

MILLS JS Jr (1972) An Application of a Decision Support System to Budgeting. SM Thesis, MIT, Cambridge, MA.

O'BRIEN RK (1972) A Mixed Integer Programming Model for Long-Range Financial Planning under Uncertainty. SM Thesis, MIT, Cambridge, MA.

PARK YC (1972) Industrial Noise Abatement Market. SM Thesis, MIT, Cambridge, MA.

PENNY MJ (1972) The Influence of Company Conditions and Practices on Motivation. SM Thesis, MIT, Cambridge, MA.

PETERSON CD (1972) Implementing Management Science: An Analysis of the Problem and a Suggested Strategy. SM Thesis, MIT, Cambridge, MA.

PICARD DA (1972) Strategic Planning of the American Multinational Firm in France. SM Thesis, MIT, Cambridge, MA.

PUTZEL R (1970) A Gestalt View of Organization Development. SM Thesis, MIT, Cambridge, MA.

DE QUADROS AC Jr (1970) Response Surface Theory: An Optimization and Prediction Technique. SM Thesis, MIT, Cambridge, MA.

REARDON DM (1972) A Study of the Cable Television Industry. SM Thesis, MIT, Cambridge, MA.

ROLLINS RH II (1970) Closing of the NASA Electronics Research Center—A Study of the Reallocation of Space Program Talent. SM Thesis, MIT, Cambridge, MA.

SALOOM JA III (1972) Organized Labor and Adjustment Assistance. SM Thesis, MIT, Cambridge, MA.

SANDBERG GG (1972) The Organization of Multinational Corporations. SM Thesis, MIT, Cambridge, MA.

SAVOLDI JA III (1970) The Imported Car Market: Problem or Opportunity for the United States Manufacturer? SM Thesis, MIT, Cambridge, MA.

SCHLARB WP (1970) An Estimation of Supply Allocation Among User Classes of Natural Gas Utilities. SM Thesis, MIT, Cambridge, MA.

SCHOMAN KE Jr (1970) Selection of a Major Course of Study by MIT Freshman. SM Thesis, MIT, Cambridge, MA.

SCHROEDER TE (1972) A Quantitative Framework for the Analysis of Business Strategy with an Application to the New England Footwear Industry. SM Thesis, MIT, Cambridge, MA.

SCHWARTZ SJ (1972) Political Campaign Funding Proposals: An Analysis. SM Thesis, MIT, Cambridge, MA.

SEDGWICK DA (1972) The Modern Volunteer Army—An Approach Toward Attainment. SM Thesis, MIT, Cambridge, MA.

SEYMOUR WE (1972) The Industrial Relations Climate in Great Britain: A Deterrent to Investment? SM Thesis, MIT, Cambridge, MA.

SHEEHAN EJ (1972) A Breath of Fresh Air: A Study of the Organizational Structure of the Environmental Protection Agency. SM Thesis, MIT, Cambridge, MA.

SHUMATE CM (1972) Design and Implementation of a Planning Process in a Community Action Agency: An Organization Development Approach. SM Thesis, MIT, Cambridge, MA.

SIDES R (1972) An Inventory Model for Style Goods. SM Thesis, MIT, Cambridge, MA.

SILVER A (1970) An Algorithm for Adjusting Production Rate as a Function of Assembly Line Balancing Costs. SM Thesis, MIT, Cambridge, MA.

SMITH JB III (1970) Seasonal Components in Security Returns. SM Thesis, MIT, Cambridge, MA.

SOSA DE BOUTAUD AA (1970) Machine Substitution in an Abrasive Job Shop. SM Thesis, MIT, Cambridge, MA.

SPENCE WF (1972) An Examination of the Nature, Development and Structure of the Market for Commercial Paper During the Period from 1965 to 1970. SM Thesis, MIT, Cambridge, MA.

STEINECKE C III (1970) Use of Marketing Research in Editorial Management. SM Thesis, MIT, Cambridge, MA.

STAFFARONI JM (1972) An Analysis of School District Regionalization in Massachusetts. SM Thesis, MIT, Cambridge, MA.

STONE GE (1972) Towards a Greater Understanding of Neighborhood Health Centers. SM Thesis, MIT, Cambridge, MA.

THEIS GF (1970) Performance of a Functional Organization While Undertaking a Secondary Task. SM Thesis, MIT, Cambridge, MA.

TRELEWICZ DB (1970) The Influence of Acquisitions on the Stock Prices of Conglomerate Companies. SM Thesis, MIT, Cambridge, MA.

UPSON JF (1970) Industrial Investment in North-East Brazil After the First Ten Years of Sudene (1959–1969). SM Thesis, MIT, Cambridge, MA.

URBAN LJ (1970) Practice and Research of R & D Management. SM Thesis, MIT, Cambridge, MA.

VANDER DUSSEN NR (1970) Corporate Growth Patterns. SM Thesis, MIT, Cambridge, MA.

WAGNER JH II (1970) The Impact of an Interactive Risk Analysis Model on Decisions Under Uncertainty. SM Thesis, MIT, Cambridge, MA.

WEINER H (1970) Role Perception and Organizational Ineffectiveness in the Foreign Service. SM Thesis, MIT, Cambridge, MA.

WELLS RA Jr (1970) Factors Which Motivate White Collar Employees to Organize Collective Bargaining Units. SM Thesis, MIT, Cambridge, MA.

WHITE DH (1970) The Decision to Undertake Large Computer-Based Management Systems Developments. SM Thesis, MIT, Cambridge, MA.

WIDMER H (1970) Marginal Costs of Nuclear Energy. SM Thesis, MIT, Cambridge, MA.

WILSON DM (1970) A Conceptual Model for Managing the Architectural Programming Process at a University. SM Thesis, MIT, Cambridge, MA.

WINTER CR (1970) The Economics of Petroleum Exploration in the Middle East. SM Thesis, MIT, Cambridge, MA.

WOODBURY MG (1970) Experiments in Restructuring the Property Tax. SM Thesis, MIT, Cambridge, MA.

YECHESKEL E (1970) Cost Model for Scheduling Mix Change in the Automobile Industry. SM Thesis, MIT, Cambridge, MA.

ZARAGOŻA-CALVET CE (1970) An Interactive Model for Allocating Sales Effort Across the Product Line.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 469-478. Pergamon Press Ltd 1978. Printed in Great Britain

EDITORIAL
The Vroom-Yetton Leadership Model

TALK TO senior managers about what they perceive are important ingredients in the managerial process, and sooner or later they come round to talking about leadership. For example, "Leaders are born, not bred", is a favourite epitaph with which to conclude a sermon on the shortcomings of business schools. You can design educational programmes to pump in knowledge about facts and figures, to transmit an understanding of the complexities of business life and society, to simulate problems for which solutions have to be found, to appreciate the implications of conflict and power in industrial systems. But management involves more than all that; it involves the elusive quality of leadership, it requires determination and foresight, the ability not only to solve a problem but to sell it, not only to persuade your subordinates but to inspire in them the confidence to follow you. What you cannot do in a business school is to implant this quality of leadership, and without it you cannot teach a student how to become a manager.

Or can you? Leaving aside the attributes of age and experience which—as in any other trade or profession—managers find useful in the execution of their tasks, is it not possible for latent talents to be nurtured through a structured course of study? While no business school would claim to be able to change the personalities of their students (or to want to, for that matter), its educational programme cannot be merely regarded as a knowledge pumping process; it provides a framework within which facts, techniques, problems and stimuli help a student to mature, to develop attitudes and to mould his character. Does all this add up to the emergence of a quality of leadership?

It all depends, of course, on what you mean by "leadership". Discussions about business schools' curricula, selection and recruiting procedures, evaluation of managerial performance, staff promotion policies—they can all be reduced to semantic quibbles when "leadership" is identified as a sought-after characteristic. It is not surprising, therefore, that, from time to time, brave attempts are made to delineate this concept, to analyse and categorize it, even to measure it. Of these, perhaps the bravest, the most detailed and the most ambitious is the Vroom–Yetton Model [2], which I now propose to discuss in some detail.

But before I turn to the model itself, a few preliminary comments may be appropriate by way of an introduction. The natural reaction to the question "What do you mean by leadership?" is to start with an explicit definition, such as the one given in a dictionary (the Concise Oxford Dictionary, 1964, defined the verb *lead*, among other definitions, as "to guide actions or opinions of, bring by argument etc. to conclusion, induce to do") or refer to synonyms and antonyms as an oblique definition (according to Webster's New Dictionary of Synonyms, 1973, the synonyms of *leader* are head, chief, chieftain, master, and the antonyms disciple, adherent, henchman, satellite, partisan, follower). While such definitions give a reasonably good "feel" for the term "leadership" in its general use, they are inadequate in a technical context, in that they fail to tell us the degree to which leadership is exercised in a particular situation, or how two leaders are to be compared (except in general qualitative terms).

The alternative is to turn to the *manifestation* of leadership. We can look at the behaviour of the followers and try to infer what traits of leadership have caused, or contributed to, observed behaviour characteristics, or better still, we can analyze the leader's managerial style and determine which of a given pre-defined set of styles fits most closely a particular case. Assuming that such a set can be properly categorized, this latter method (of directing our attention to managerial style rather than using inference based on the followers' pattern of behaviour) is the more direct and the easier to develop.

If we consider the main causes that determine why a manager behaves in a particular way, perhaps four main modes may be outlined:

(a) Personalistic

(b) Constitutionally determined

(c) Recipient dependent

(d) Problem oriented

The first is the *personalistic* mode, which is entirely a function of the personality of the manager. If he is autocratic by nature, he is then assumed to behave autocratically whenever decisions are called for, irrespective of how his colleagues and subordinates feel about it, and irrespective of the characteristics of the problems that need to be solved. The second possible mode is *constitutionally determined*, in that certain procedures and even rituals have to be observed before decisions are decreed. Such rules of behaviour may have evolved over a period of time, or may have been laid down by a previous leader; they form a kind of constitution, written or unwritten, within which a new leader is expected to assume his role, and he

Vol. 6
1978

must learn the basic rules of the game before he can play. The third mode might be *recipient dependent*, where the manager adapts his style to suit the group of subordinates who will be most reactive to a particular decision, so that while he may behave autocratically towards one group, his style in dealing with another may be consultative and accommodating. The fourth mode ignores the personal characteristics of the decision maker, or the possible response of the subordinates, and concentrates on the problems for which decisions are sought; the *problem oriented* approach envisages the manager adapting his style to meet the needs of the particular problem that he has to face, so that a given group of subordinates will experience, in the course of time, different managerial styles exercised by their leader.

This is, of course, an oversimplified thesis. For a start, the four modes cannot be expected to fall neatly into self-contained compartments: it is unlikely that personalistic behaviour can, in practice, be totally immune from the effects of the many external forces that impinge on the decision maker; and even in the most rigidly defined constitutional environment, the personality of the leader often asserts itself to some degree. Similarly, the environment cannot be regarded as entirely separate from the groups of recipients that form a part of it, since they often contribute to the development of the constitutional framework, and equally the type of problems to be solved may well be correlated with the kind of environment and the level of bureaucratic rigidity from which they emerge. Furthermore, a manager with a certain personality may be more attracted to, and therefore may be more likely to be found in one environment with its own characteristic problems, than in another, and the same may apply to groups of subordinates and their level of tolerance towards the various managerial styles of their leader.

Thus, a model of managerial behaviour based exclusively on the four modes listed above is not likely to take us very far, but they are perhaps a useful starting point, in that the first two are problem independent, whereas the latter two comprise what may be called "situational variables" [2]. The first question, therefore, that we need to consider is whether managerial behaviour is found in practice to be largely personalistic (and/or constitutionally determined) or a function of situational variables. Yetton and Vroom have no doubts about that. They point out that students of managerial psychology have been "nurtured on a diet of Theory X and Theory Y, and other individual difference models of leadership" [2, p. 133], in which a manager is assumed to adopt a particular dominant style and stick to it, with deviations from his chosen style regarded as a normal random error. But the authors' own research has led them to conclude that "the evidence overwhelmingly supports the thesis of problem characteristics as determinants of leadership style" [2, p. 145] and that the trait theory

TABLE 1. VROOM AND YETTON'S TAXONOMY OF MANAGERIAL STYLES [2, p. 134]

AI	You solve the problem or make the decision yourself, using the information available to you at the present time.
AII	You obtain any necessary information from subordinates, then decide on a solution to the problem yourself. You may or may not tell subordinates the purpose of your questions or give information about the problem or decision you are working on. The input provided by them is clearly in response to your request for specific information. They do not play a role in the definition of the problem or in generating or evaluating alternative solutions.
CI	You share the problem with the relevant subordinates individually, getting their ideas and suggestions without bringing them together as a group. Then *you* make the decision. This decision may or may not reflect your subordinates' influence.
CII	You share the problem with your subordinates in a group meeting. In this meeting you obtain their ideas and suggestions. Then, *you* make the decision which may or may not reflect your subordinates' influence.
GII	You share the problem with your subordinates as a group. Together you generate and evaluate alternatives and attempt to reach agreement (consensus) on a solution. Your role is much like that of chairman, coordinating the discussion, keeping it focused on the problem and making sure that the critical issues are discussed. You can provide the group with information or ideas that you have but you do not try to "press" them to adopt "your" solution and are willing to accept and implement any solution which has the support of the entire group.

should therefore be abandoned in favour of a contingency theory; thus, "whereas once there were only participative or autocratic managers, we now find that it makes as much sense to talk about participative and autocratic problems" [2, p. 146].

We, therefore, need a more detailed classification of managerial styles, and the Vroom–Yetton theory identifies five distinct styles (shown in Table 1), ranging from the completely autocratic (AI) to group decision making (GII), where the manager acts as an unbiased chairman looking for a consensus. Whether five is the right number for such a classification is an open question, though perhaps not a crucial one: clearly, the more categories there are, the finer are the nuances between adjacent categories, and the more difficult it becomes to assign a single category to a given practical situation. Even with the five proposed in Table 1, where the differences between them are quite distinct, attribution is not always obvious. What happens, for example, when you share the problem with some relevant subordinates but not with others? Is it AII or CI? Or when you share it with subordinates and have a group meeting with some? Is it CI or CII? And what precisely is meant by "sharing a problem"? There are so many shades of openness in discussion and so many nuances of interpretation. And what about your relationships with your peers? Do they not have some influence on how you behave? But there is no denying that Vroom and Yetton's descriptions of the five styles are a great improvement on the restrictive Theory X and Theory Y classification,[1] in that they allow a wider range of alternatives to be observed and tested.

[1] Not even McGregor himself asserted that these "cosmologies", as he called them, exist to the exclusion of others [1, p. 33].

Vol. 6
1978

Having proposed their five managerial styles, Vroom and Yetton turn their attention to two models of behaviour, *the normative model* (which purports to tell managers how to act) and *the descriptive model* (which seeks to determine how in fact they do behave in practice). In either case, it is necessary to define a set of situational variables to which we then attempt to relate either the prescriptions in the normative model or the empirical findings in the descriptive model.

The seven situational variables proposed by Vroom and Yetton are listed at the top of Fig. 1 (the first three pertain to the problem for which a decision is sought and the other four to acceptance of the solution by subordinates), from which they proceed to map out their normative model in the form of a decision tree. The twelve terminal nodes of the tree (on the right) show the set of feasible managerial

A. Does the problem possess a quality requirement?
B. Do I have sufficient information to make a high-quality decision?
C. Is the problem structured?
D. Is acceptance of the decision by subordinates important for effective implementation?
E. If I were to make the decision by myself, am I reasonably certain that it would be accepted by my subordinates?
F. Do subordinates share the organizational goals to be attained in solving this problem?
G. Is conflict among subordinates likely in preferred solutions?

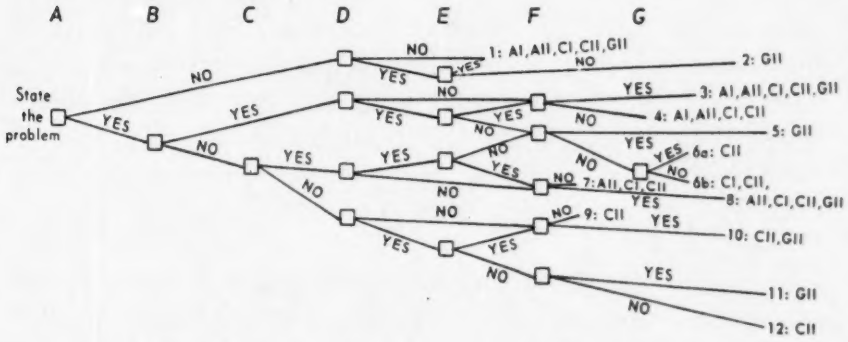


FIG. 1. *The Vroom-Yetton normative model* [2, p. 138].

styles appropriate to the respective nodes: for example, node 1 is arrived at when questions A and D are answered in the negative, in which case all five styles remain feasible, whereas node 2 is obtained when the answers to questions A, D and E are "yes", "no" and "yes" respectively and then only style GII is feasible.

I have to confess that the methodology associated with the seven situational variables makes me more than a trifle uneasy. Not only is the expected response to each of the seven questions confined to a "yes" or "no" answer, but most of them remind me of some of the incomprehensible questionnaires that I keep getting in the post. "Does the problem possess a quality requirement?"— is this the same as "Is the problem important?" and if so, how do I measure its importance? If the problem is quite trivial, in that it has no significance to the well-being of the enterprise or its members, what manner of problem is it and why does it merit the attention of the manager?

And if by virtue of its being a problem worthy of managerial attention it automatically possesses a "quality requirement", then why list this question as a situational variable?

Or take question B: "Do I have sufficient information to make a high quality decision?" What is "sufficient" and what is "high quality"? And in question C, "Is the problem structured?", how do I determine whether a problem is structured? Does structure relate to the degree of frequency and replication with which the problem occurs, in other words with having to apply programmed or non-programmed decisions (using Simon's terminology) to solve it? And what about all those problems that start as unstructured and end up as structured, and those that seem structured to begin with but become increasingly ill-defined as the manager continues to probe into objectives and constraints?

Similar queries arise for all the other situational variables and I remain sceptical about the validity of a methodology that seeks "yes" or "no" answers to issues which are inherently complex and pose fundamental questions about evaluation and measurement. Furthermore, in the various validation studies referred to by Yetton and Vroom the managers did their own coding and interpretation (in many cases they even chose the problems to be recorded and analysed), thereby introducing yet another source of bias and inconsistency since different managers may well judge given situations differently; admittedly, Yetton and Vroom recognize this shortcoming, but in defence they "report that errors in judgement about the situational characteristics are fairly low" [2, p. 147]. In view of what was said earlier, I have my doubts.

To the many difficulties already alluded to, it is relevant to point out another major weakness in the scheme of situational variables, namely that it is a presentation of *static* problems that face the decision maker. We are supposed to imagine him as having suddenly to cope with a problem, which he proceeds to analyse by the seven situational variables and then to solve it (if he follows the normative approach in Fig. 1) by adopting one of the managerial styles listed in Table 1. The problem has neither a past nor a future (except in terms of the quality of the outcome), it does not evolve, it does not change in content or significance. It is a mere discrete event, a single stimulus in search of a single response. It is not only a static, but a mechanistic model of decision making. And because of the supposedly static nature of the problems that the manager needs to solve, it is assumed that we can ask him, even long after the event, to analyse and codify them, and that he can do so with complete recall, without error and without bias.

Reality, however, is very different. Not only do problems change as they are being processed, but one's view of problems changes with time. What seems important and urgent today may sink into oblivion

compared with the challenges facing the manager tomorrow. What seems like total acceptance by subordinates one day may be perceived very differently when subsequent repercussions begin to materialize.

If I had to summarize a set of responses to the seven situational variables in Fig. 1, it would go something like this:

- every problem has a quality requirement, otherwise it does not deserve my attention;
- I never have sufficient information, but (time and cost constraints being what they are) I have to make do with what I have;
- no routine and highly structured programmable problem should remain unprogrammed, and if for reasons of human relations there is a need to keep managerial staff involved, then it should be delegated whenever possible;
- except in trivial situations, the acceptance of decisions by subordinates, and peers, is always necessary, or at least desirable;
- involving others in decisions helps to make people committed, but I cannot expect to be absolved from the responsibility and accountability for decisions made by me, or on my behalf;
- subordinates have an interest in the survival of the enterprise in which they work, but their own goals are often at variance with those of the organization with respect to particular issues;
- there is always a conflict of interests between individual subordinates, as well as between groups.

Now, Yetton and Vroom may argue that with such a set of answers to the seven questions, I am not the kind of person for whom the coding of situational variables in any given case, nor the proposed normative approach in Fig. 1, can be of much use. I wonder, though, whether these answers, or at least most of them, are not fairly typical of a large section of the managerial population. It is possible, of course, that I and many like me will give these answers because of the particular positions and responsibilities that we hold, and which involve problems with certain common characteristics. In that case our attitudes are dominated either by our function or hierarchical level,[2] in other words, our behaviour is largely fashioned by the roles that we have to assume, so that we are back to the constitutionally determined mode of managerial style, mentioned earlier.

[2] Jago and Vroom conclude from their studies that participation, for example, "typically increases with hierarchical level" [2, p. 145].

There is another aspect of the construction of the normative model that needs to be highlighted. Before we can prescribe how a manager should behave, it is first necessary to specify a criterion by which one alternative is judged to be superior to another; then we need to compare the results of the model with empirical evidence. Yetton and Vroom do both. In order to determine the feasible set of styles in each case, they proceed to devise a set of rules based on the seven situational variables (for example, *the leader information role* states that "if the quality of the decision is important and the leader does not possess enough information or expertise to solve the problem by himself, then AI is eliminated from the feasible set", *the goal congruence rule* demands that "if the quality of the decision is important and subordinates are not likely to pursue the organization goals in their efforts to solve this problem then CII is eliminated from the feasible set", and so on), from which the normative tree in Fig. 1 can then be constructed. The guiding principle is that of a cost/benefit analysis; as Yetton and Vroom put it:

> "Now any increase in participation implies a cost in terms of the additional manpower resources it consumes. Assuming that the managerial manpower consumed by decision-making increases monotonically with an increase in participation, Yetton and Vroom define the optimal short-term managerial style as the minimum level of participation (minimum cost) which would generate a decision that would satisfy the quality and acceptance needs in the situation. This is called the minimum cost strategy." [2, p. 137]

The blind faith that the authors have in the powers of such a cost/benefit analysis is indeed admirable but misguided, since the assignments of costs in situations of this kind are notoriously arbitrary and unreliable. What is clearly a fixed cost element to one manager may be regarded as an opportunity cost by another, and the arithmetical determination of a single cost of a meeting of a group of people to discuss a particular problem is often no more than a hypothetical cost saving if that meeting were not held. This is not to say that communication and participation do not absorb resources and cost money; of course they do. But the assignment of numerical cost values to such discrete events is fraught with so many difficulties, that it is questionable whether they can serve as the sole criterion for differentiating between alternatives.

So, if cost minimization is not the answer, perhaps we can turn to the quality of the decision as a possible guideline. This would require managers to look back at their experience and code problems according to their situational variables, to identify the particular managerial style adopted in each case and then to rate the quality of the solution on a scale (a 7-point scale was used by Vroom and

Yetton in one study and an 11-point scale by Yetton in another [2, p. 139]). Armed with these data, so the argument goes, the investigators can proceed to identify the styles pertaining to those combinations of situational variables which are associated with high quality decisions. The road to a normative behaviour model then becomes clear.

But the objections to this procedure are obvious. Since the scheme of situational variables is a vital component in this model building process, it follows that without it the whole structure falls to the ground, and I have already pointed to the questionable validity of the assumptions on which the seven situational variables are based. In addition, the evaluation of the quality of a decision by managers must rank as a rather dubious exercise. Apart from the inevitable bias and self-justification that are likely to be manifest, there is, once again, the effect of the time element to be considered. Only trivial and unimportant decisions can be summarily evaluated soon after the event; for significant decisions, in terms of their possible consequences to the enterprise, long periods of time must elapse before a proper analysis and evaluation can take place, the initial perception of the original problem may, by then, have become distorted, and the success or failure of a particular decision may well have been governed by external forces completely unrelated to the internal situational variables.

So much for the development of the normative model. Then Yetton and Vroom continue with a series of studies to compare the results of the model with empirical data on managerial behaviour as an exercise which is intended to validate the model. In fact, it cannot do anything of the kind. It is possible for a model that tells people how to behave to be correct and yet for people to behave differently, and conversely it is possible for people to behave as they are urged to and yet for the normative prescription to be false. A comparison of a normative model with a descriptive one can neither add to nor detract from the validity of the former; its purpose can only be to determine the discrepancies between the two and then to speculate on the reasons that cause a divergence. Even in the extreme case when the results of the two models are always identical (and when it can consequently be argued that from a statistical point of view this cannot be a coincidence), we have no proof that the normative model is correct, only that it is redundant.

Thus Yetton and Vroom's contribution to the leadership debate lies mainly in their proposition that managerial styles are not polarized between the good Theory Y believers and the bad Theory X adherents, that managers do not necessarily stick to a single style dictated by their personalities and systems of beliefs, and that task and problem orientation is an essential element to consider in empirical investigations of managerial styles. These are important contribu-

tions, and their effect on current thinking on the concept of leadership should not be underestimated. But all this is a far cry from accepting Yetton and Vroom's thesis that we are now able to codify and quantify the variables that determine or even influence styles of leadership adopted by managers, let alone assert that we have proven prescriptions to tell them how to behave.

SAMUEL EILON
Chief Editor

REFERENCES

1. EILON S (1977) *Aspects of Management*. Pergamon Press, Oxford.
2. YETTON PW & VROOM VH (1978) The Vroom–Yetton model of leadership—an overview, in *Managerial Control and Organizational Democracy* (Eds KING B, STRENFERT S, and FIEDLER FE), pp 133–149. John Wiley.

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 479–484
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0479$02.00/0

Optimal Solutions Versus 'Good' Solutions: An Analysis of Heuristic Decision Making

JACK A FULLER

University of Northern Iowa, USA

(*Received April* 1978; *in revised form June* 1978)

The portion of business problems which lend themselves to solution by existing optimization techniques are a minor subset of the business problems which can be defined and attacked by logical devices. Heuristics simplify the environment of the decision maker by permitting him to make decisions quickly without considering the innumerable ways that each decision could be made. This paper compares the heuristic programming approach to more traditional operations research techniques of decision making.

INTRODUCTION

IN THE typical complex business environment, the problems being faced are orders of magnitude more difficult than those encountered in the scientific community. The advantage of using heuristic rules is that they limit the search by reducing the number of alternatives to a manageable size [13]. The heuristic approach to typical problems is essentially a simulation of how a human might search for a good solution to each type of problem if he had precise rules for evaluating alternatives [10]. Because of the great speed of electronic computers, such heuristic programs can identify and evaluate alternatives quickly and accurately.

Actually, heuristics are used on a daily basis by almost everyone and can be learned from experience. Often a heuristic may be developed and followed without the user actually being aware he is using it. Michael [10] presented the following examples of everyday heuristics:

In golf: Use an old golfball on a hole with a water hazard.

In driving: When traffic backs up on Eden's expressway, take Foster Avenue to Lake Shore Drive to get to the loop.

The following are some of his examples of simple heuristics found in a variety of functional areas in business.

In management: Handle the exceptional problems, not the routine.

In finance: Keep current ratio greater than two to one.

In the stockmarket: Don't buy before a long holiday weekend.

In inventory control: Re-order when a storage bin is half full.

HEURISTIC PROGRAMMING APPROACH

Heuristic programming is particularly suited to two main types of problems: (1) those that are too large to manipulate with traditional operations research models and (2) those that are too ill-defined or loosely structured to be expressed in the mathematical relationships required of the traditional operations research models. In the first case, some problems are simply too large for the formal operations research algorithms and/or even the largest computers. The problem in this case is not only size of the problem but also that the search phase of the decision making process is just

too extensive. For example, in a complete warehouse location problem there are so many location evaluations that the optimal solution is often computationally impractical. The heuristic approach can aid such a situation since it basically is a shortcut through the search phase of such decision making problems. This characteristic of heuristic decisions makes heuristic programming especially adaptable to problems that are very large and complex, that is, problems where the search phase and decision making could be extensive.

In the second case, problems may be too loosely structured for the formal algorithms of the traditional operations research models. Such formal operations research algorithms demand very specific inputs and clearly stated and relatively simple objective functions. In contrast, much of the decision making that a typical manager does involves situations that are very loosely structured. As Michael [10] states, the solution and even the exact definition of such problems will vary, depending on the manager's judgment, creativity, and prior experience. In such cases, heuristic programming can yield a solution that, although not optimal, is usually satisfactory and better than intuitive decisions. In such loosely structured situations, the manager may be looking for satisfactory answers, rather than optimizing. In such cases the heuristic programming solution is entirely acceptable and desirable.

Newell [12] identified and classified the following five basic methods of heuristic programming: (1) generate-and-test, (2) hill-climbing, (3) match, (4) heuristic search, and (5) induction. He suggested that others existed; however, they are extremely difficult to classify since they are, in many cases, combinations of the previously mentioned methods and do not occur as complete methods by themselves.

In the generate-and-test method, all that is necessary is a generator of possible candidates for a solution and a test of the candidates to find if they are indeed solutions. Opening a combination lock without knowing the combination is an example of such a generate-and-test method. Match method, on the other hand, is more efficient than the generate-and-test method. The key to using the match method is to define a set of possible solutions to a form with variables. In this case, one limits the search for a solution to the size of the form rather than to the size of the set of all possible solutions, as was the case in generate-and-test.

The hill-climbing method is one of the heuristic programming procedures for seeking an optimum. In hill-climbing, possible solutions are generated. A solution is then compared against the best solution generated previously and then is either discarded or, if it is a better solution, it replaces the previously retained solution. Even though the hill-climbing procedure is concerned with seeking an optimum solution, it has no procedure for guaranteeing that it will be found or even identifying one when it is found. Rather, hill-climbing yields local peaks which may or may not be the highest peak in the feasible set of solutions.

Newell [12] refers to induction as being a classical reasoning process. For use in heuristic programming, two features of induction are essential. First, a problem must be cast as one of finding a function or mapping of the given data into associated or predicted data. Second, the method uses a form or kernel for the function which can be matched in the sense of the matching method of induction.

There are several common features associated with the above mentioned five heuristic programming methods. First, the assumptions of the methods need not be strong. Second, the methods allow flexibility and creative combinations of pure methods without necessarily treating each method as a closed subroutine with its output being tied to the input needed by the next method. Last, the heuristic methods make much weaker demands for information than the more structured operations research techniques. This characteristic allows for a great deal of generality in the heuristic methods.

A HEURISTIC DECISION RULE APPLICATION

As an example, three heuristic decision rules applicable to scheduling production and work force in an aggregate production planning framework were developed, tested, and compared. According to Newell [12] these heuristic decision rules would be classified as hill-climbing heuristic programming methods. In this case, possible solutions were generated and compared against those previously generated, so that the best solution could be identified.

The application of this set of heuristic decision rules should better help exemplify what was previously discussed with respect to heuristic decision making.

Using appropriate simulation techniques, the heuristic decision rules were tested on a variety of demand patterns and forecasts. The basic procedure was to choose a heuristic rule and a demand pattern. For each demand pattern, several forecasts were used in testing the decision rule. For a particular forecast of a demand pattern, the heuristic decision rule was simulated. The appropriate variables for the heuristic rule chosen were varied, and an assortment of production schedules resulted. The computer programs were designed to select that production schedule for a particular program which would result in the least comparable cost. This procedure was then repeated for the remaining forecasts of the demand pattern and then similarly for the remaining demand patterns and their forecasts.

The result of the simulation of one heuristic decision rule was several production schedules, one schedule for each of the forecasts of each of the different demand patterns. Each schedule had associated with it particular values of the total production costs, related values of the variables in the heuristic decision rule, and the resultant planned production levels for each period in the study. With this information it was possible to identify that forecast and those heuristic decision rule variable values that perform best as measured by least total comparable production cost for a particular historical demand pattern.

DESCRIPTION OF PROPOSED HEURISTICS

The first heuristic decision rule, hereafter referred to as Rule 1, took the form

$$P_t = \bar{D}_t + (D_{t-1} - \bar{D}_{t-1}) + (S - I_{t-1}) \quad (1)$$

where

P_t = planned production level in period t,
$\bar{D}_t$ = forecasted demand level in period t,
D_{t-1} = actual demand level in period $t-1$,
$\bar{D}_{t-1}$ = forecasted demand level in period $t-1$,
S = desired safety inventory stock level,
I_{t-1} = actual inventory level at the end of period $t-1$, and
α = a smoothing constant (not used in Rule 1).

The safety inventory stock level was allowed to vary while using a given forecasted demand pattern. Such a procedure gave various possible production levels for each planning period in the planning horizon. The procedure was to pick a value for S which then remained constant as the rule was used throughout the planning horizon. Successive values of S were chosen, and the procedure was repeated until the selected range of values was exhausted. The resulting production schedules were compared on a cost or profit basis, and the most appropriate safety inventory stock levels for the demand patterns were chosen.

The rule allowed production to follow the forecasted demand levels unless there had occurred in the previous period either deviations between forecasted and actual demands or between actual and desired inventory levels. Rule 1 allowed for variations in production to account for such deviations. At the end of a particular time period, the actual demand and inventory levels were known, and production could then be scheduled for the next period.

Heuristic Rule 2 took the form

$$P_t = \bar{D}_t + \alpha (D_{t-1} - \bar{D}_{t-1}) + (S - I_{t-1}) \quad (2)$$

Before this rule was used, values for S and α were chosen. These values then remained constant during the planning horizon as a production level was obtained for each period in the planning horizon for the forecasted demand pattern being used. A given value of desired safety stock inventory level and alpha resulted in a production schedule. Other production schedules were obtained by using different values of S and α for the same forecasted demand pattern. The production schedules were compared on a cost or profit basis to determine the most appropriate values for the safety inventory stock and smoothing constant.

Heuristic Rule 3 took the form

$$P_t = \bar{D}_t + \alpha (D_{t-1} - \bar{D}_{t-1} + S - I_{t-1}) \quad (3)$$

This rule was similar to (2), the difference being that here only some fraction of the difference between the desired and actual inventory level in the previous period influences production in the present period. This should decrease some of the large variations in production levels which may be evident in Rules 1 and 2.

SIMULATION OF HEURISTICS

Using appropriate simulation techniques, an IBM 360 Model 50 computer and the FORTRAN programming language, the three heuristic decision rules were tested using three demand patterns (details given in Fuller [4]). The computer programs were designed to select that production schedule for a particular program which would result in the least "comparable cost." Comparable cost is here defined as the total production cost during the time span under consideration plus an additional cost for lost profits and lost contributions to fixed costs when stock-outs or lost sales occurred minus the production costs of the units left in inventory at the end of the total time span [4].

The result of the simulation of one heuristic decision rule was nine production schedules (one schedule for each of the three forecasts of each of the three different demand patterns). Each shedule had associated with it particular values of total production costs, total comparable production costs, related values of the variables in the heuristic decision rule (i.e., desired safety inventory stock and the smoothing constant alpha), and resultant planned production levels for each period in the study. With this information it was possible to identify that forecast and those heuristic decision rule variable values that performed best, as measured by least total comparable production costs, for a particular historical demand pattern.

The beginning inventory and production levels used in initiating the studies were assumed in all cases to be equal to values obtained from the first demand pattern [4]. This means that the beginning inventory was 215 units of product and the production level in the period preceding the first period in each study was 130 units of product. The desired safety stock level of inventory, S, was varied in the simulation from 0 to 80 in steps of 1. The procedure was to choose a forecast and a value of S. This resulted in a production schedule from which a total production cost and a total comparable production cost was computed. This was repeated for other values of the desired safety stock level until all 81 possibilities were exhausted. The production schedule which resulted in the least comparable total production costs was retained as the most desirable production schedule for the particular forecast under consideration. In the case of Rules 2 and 3 the smoothing constant alpha was varied from 0 to 1.0 in steps of 0.1.

The actual number of simulations performed in this study was dependent on the particular heuristic rule under investigation. For Rule 1 the simulation for a particular demand pattern was performed using three forecasts and varying the safety stock S over a twelve month study. This means that Rule 1 was used in the simulation to develop 2,916 months of simulated production for a given demand pattern. Since Rules 2 and 3 also used a smoothing factor α, they each were used to develop 32,076 months of simulated production for a given demand pattern.

ANALYSIS OF SIMULATION RESULTS

The best schedule, in terms of having the least total comparable production costs over the twelve months involved in each study, was chosen from among the three available for each of the three demand patterns used in testing each of the three heuristic decision rules. These production costs are compared in Table 1. The table indicates that Rule 3 was the best to use in scheduling production and work force for the first demand pattern (i.e. actual company data), while Rule 1 was the least effective. Rule 2 was the best decision rule to use for the other two demand patterns.

CONCLUSIONS

For the demand patterns investigated, the above analysis does not appear to distinguish any single heuristic decision rule as consistently superior to the remaining rules. As seen in Table 1, the relative differences in total comparable production costs for some of the heuristic rules for given demand patterns are almost negligible. Although heuristic Rule 3 was ranked number one on only one of the three demand patterns, the percentage difference between it and the best rule on those occasions was so small as to make it appear to be a generally applicable rule for demand patterns approximating those under investigation.

Vol. 6
1978

TABLE 1. COMPARISON OF HEURISTIC RULES

Demand pattern	Heuristic rule	Safety stock (if used)	Alpha (if used)	Total production cost	Total comparable cost
one	one	34	N.A.	$1,500,490	$1,169,740
one	two	42	0.0	1,450,658	1,152,983
one	three	4	0.2	1,301,242	1,143,217
two	one	0	N.A.	2,426,248	2,426,248
two	two	1	0.7	2,425,890	2,425,890
two	three	0	1.0	2,426,248	2,426,248
three	one	1	N.A.	1,835,644	1,831,969
three	two	4	0.6	1,831,712	1,831,712
three	three	2	0.9	1,831,917	1,831,917

The relevant question, however, should not be whether the above heuristic decision rules are valid or not, since they were only presented as examples of the heuristic decision making approach to problem solving. The important question is whether heuristic programming is a valid decision making tool to be utilized by managers or management scientists in a business environment. The question arises because heuristic programming, by definition, does not contain a specific algorithm and researchers often come up with slightly different heuristics and, therefore, different solutions. As Michael [10] points out, the separation between heuristic reasoning and traditional decision making techniques must be kept clear. This should help minimize the controversy between traditional operations research techniques and heuristic programming and focus attention on the results which heuristic programming may supply.

Since heuristic programming does not guarantee optimal solutions, evaluating the results generated from heuristic programming is quite difficult. One approach would be to compare the solutions which result from the use of heuristic programming tools to that of more traditional operations research techniques (i.e. linear programming, nonlinear programming). The comparison could be made on not only solutions but also on development and computation time. This may not be possible in most situations, however, since heuristic programming is most typically applied to problems which are not easily, if at all possibly, solved by the usual operations research models. The three heuristic decision rules developed earlier were compared just among themselves. To validate these heuristic decision rules, they should be compared against more traditional operations research techniques.

Wiest [14] has suggested that a manager contemplating the use of heuristic programming should ask himself the following four questions: (1) Does it produce better results than our present methods do? (2) Are there incremental savings in resources? (3) Are computational effort and expenses reduced without sacrificing the quality of work? (4) Is the information produced more timely and are decisions reached earlier than by present methods? In essence these four questions might be summarized as the following: Is the heuristic programming approach the most realistic approach for us to take? This may oftentimes be the case, since the assumptions and simplifications which must be made by decision makers in utilizing traditional operations research techniques may make the results obtained from those techniques invalid. Whereas, although the heuristic programming approach does not guarantee optimal solutions, the reality which it brings to the decision making environment may make its solutions better than those obtained through the use of other methods.

REFERENCES

1. ARMOUR G & BUFFA E (1963) A heuristic algorithm and simulation approach to the relative location of facilities. *Mgmt Sci.* **9**, 294–309.
2. CHURCHMAN C (1967) Wicked problems. *Mgmt Sci.* **14**, 141–143.
3. CLOONAN J (1966) A heuristic approach to some sales territory problems. *Proc. of the Fourth Int. Conf. on Opl Res.*, John Wiley, New York, pp. 284–292.
4. FULLER JA (1973) Forecasting techniques as related to aggregate production planning. *1973 AIDS Proc.*, pp. 295–297.
5. GERE W (1966) Heuristics in job shop scheduling. *Mgmt Sci.* **13**, 167–190.
6. GORDON P (1962) Heuristic problem solving. *Bus. Horizons* **5**, 52.
7. HINKLE C & KUEHN A (1967) Heuristic models. *California Mgmt Rev.* **10**, 59–68.

8. HURST E & MCNAMARA A (1967) Heuristic scheduling in a woolen mill. *Mgmt Sci.* **14,** 182–203.
9. KARY R & THOMPSON G (1964) A heuristic approach to solving traveling salesman problems. *Mgmt Sci.* **10,** 225–248.
10. MICHAEL G (1972) A review of heuristic programming. *Dec. Sci.* **3,** 74–100.
11. MUDAR P (1969) Heuristic programming. *Chem. Engng Progress* **65.**
12. NEWELL A (1969) Heuristic programming: III-structured problems in *Progress in Operations Research,* Vol. III, (Ed. ARONOFSKY J), John Wiley, New York, pp. 360–414.
13. TONGE F (1961) The use of heuristic programming in management science. *Mgmt Sci.* **7.**
14. WIEST J (1966) Heuristic programs for decision making. *Harv. Bus. Rev.* 129–143.

ADDRESS FOR CORRESPONDENCE: *Dr Jack A Fuller, School of Business, University of Northern Iowa, Cedar Falls, Iowa* 50613, *U.S.A.*

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 6, pp. 485–492
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0485$02.00/0

The Pressure for Growth as an Organizational Problem

LARS ENGWALL
University of Uppsala, Sweden

(*Received June* 1978)

A basic argument of this paper is that organizations are under pressure to increase their revenues because of rising aspirations of those who provide goods and services to the organization. The paper then presents an analytical tool which can be used to identify different strategic alternatives for handling the described situation. These alternatives can be classified into three main categories: customer oriented, technology oriented and provider oriented strategies. The application of the method is illustrated by an application to the situation of the newspaper industry.

1. INTRODUCTION

IN MOST developed countries, the first twenty years after World War II were characterized by spectacular growth performance. In the 1970's, the goal of economic growth was questioned in works such as those of the Club of Rome (cf. e.g. Meadows [8]) and Hirsch [5]—the former stressing the physical limits and the latter the social limits to growth.

A situation in which there are limits to growth is a problem in itself, but that problem becomes all the more severe if, as this paper will argue, all organizations are under pressure to grow by increasing their revenues. Under such circumstances, it is of great importance to analyze the situation and to discover appropriate strategies with which to handle the pressures for growth.

This paper will provide the main arguments behind the proposition that organizations are subject to pressures for growth. It will then describe an analytical tool for identifying strategies to handle these pressures. Finally, the analysis will be applied to a particular case which concerns the newspaper industry.

2. THE PRESSURE FOR GROWTH IN ORGANIZATIONS

In order to put forward the arguments concerning the pressure for growth in organizations, it is useful to employ an organizational model like the one shown in Fig. 1. It is mainly based on works by scholars such as Barnard, Cyert, Hall, March and Simon [1, 2, 4 and 7]. The core of this model is *personnel*, i.e. those who by employing the *technology* of the organization, operate on the input provided by *external providers* (financiers, governments, and suppliers), in order to produce an output for *customers*. For this output, customers pay a price, and the revenues obtained are then distributed among providers, i.e. among personnel, financiers, governments, and suppliers. The organization is thus considered as a place of exchange of physical goods and financial resources.

The exchange mentioned takes place between two main groups, which have been referred to as customers and providers. For both of these groups, the organization is subject to competition from *competitors*, who may offer better conditions.

When discussing competition, it is usual to concentrate on the effects of competition on customer behavior. Here, however, we will stress the effects on provider behavior. We will argue that the providers in an organization will compare their conditions with the conditions of their counterparts in other organizations. In so doing, they may be likely to find at least one organization where the conditions are better. As a result, we can expect them to strive

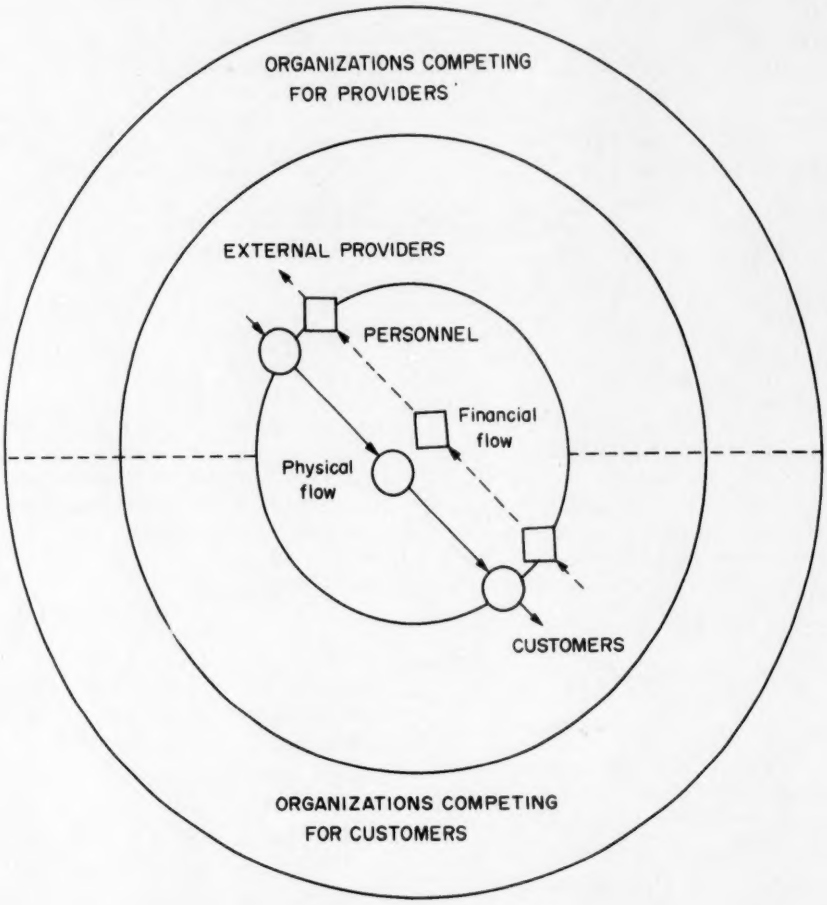


FIG. 1. *The organization model.*

to improve their conditions, i.e. to increase their compensation from the organization.

In some cases the organization may exhibit a better performance than other comparable organizations, but even this situation cannot be expected to save the organization from pressure for growth. In this case we may expect the earlier good performance to have raised the aspirations of the providers, and led them to expect a certain growth in compensation.

The above arguments are based only on the assumption that individuals prefer a higher compensation to a lower one. In their extreme form in economic theory, such arguments would imply that providers would leave the organization for another one as soon as they found another organization offering better conditions (the use of the *exit alternative* in the terminology of Hirschman [6]). In a less extreme form we would introduce inertia into the system, i.e. people employed do not want to change jobs due to the disadvantages of moving, suppliers like to keep permanent relationships for stability reasons, etc. In that case we would expect providers to be less apt to leave the organization but instead to use the *voice alternative* (cf. again [6]), i.e. to express their demands internally.

The demands for increased compensation from individual providers will add up to a pressure for growth for the organization. This pressure is enforced by the fact that all organizations are subject to the same type of pressure, and that they want to transfer this pressure to other organizations. In other words, the providers of an organization have their own providers, who have their own providers, etc., all wanting to improve their conditions.

So far the arguments have been limited to the case of monetary rewards to providers, but pressures for growth in revenues will also arise through the wish for 'empire building' on the

Vol. 6
1978

STRATEGY ORIENTATION	AIM OF STRATEGY	WITHIN EXISTING SYSTEM	CHANGE IN THE SYSTEM
CUSTOMERS	To increase revenues	1. Price hikes 2. Increased sales to existing customers	1. New customers nationally 2. New customers abroad
TECHNOLOGY	To change the transformation process	Increased efficiency of present technology	Radically new technology
PROVIDERS	To decrease the demands on the organization	1. Reducing input prices 2. Reduce input per output	1. Low wage labour 2. Cheap imports 3. Production abroad 4. Government subsidies

FIG. 2. *Different strategies to handle organizational pressure for growth.*

part of different individuals in the organization. They want to improve the status of their group, their department, etc. All this requires increasing economic resources.

3. STRATEGIES TO HANDLE ORGANIZATIONAL PRESSURE FOR GROWTH

If an organization is subject to pressure from its providers to increase their compensation, the handling of these pressures will become of vital importance. This is the case, since we may expect that different members of the organization will withdraw (use *the exit alternative*) if they are not handled appropriately. This in turn may ultimately lead to the death of the organization.

In order to analyze this situation of pressure for revenue growth the model described in the previous section can be used as a basis for an analytical tool. According to this model decision makers have three 'pure strategies' to handle pressures for growth, i.e. strategies in relation to (1) *customers*, (2) *technology*, and (3) *providers*. Within these we may make a distinction between actions (a) within the existing system, and (b) which imply changes in the system. This leads us to the matrix shown in Fig. 2 which also shows examples of different substrategies within each cell.

The easiest situation is, of course, the one in which the customer oriented strategy can be applied, i.e. when the pressure for growth in revenues can be transferred to customers. In that situation we may identify two substrategies within the existing system: 1. price hikes, and 2. increased sales to existing customers. We meet both of these strategies in our role as consumers in the form of higher prices and through promotion activities. The latter are of course also used to attract new customer groups nationally, and as national. markets approach saturation, companies often extend their promotion activities to international markets.

In many cases, which seem quite apparent today, however, the customer oriented strategy will be difficult to pursue. An alternative is then to try to change input–output relation-

ships. This can be done either within the existing system by increasing the efficiency of the present technology or by finding a new technology.

Yet another alternative is to employ the provider oriented strategy, i.e. to decrease the demands on the organization. When this is done within the existing system it is a question of negotiating with existing providers concerning the prices for their inputs and/or reducing the necessary input per output. The alternative outside the existing system here is to find new providers who are willing to provide inputs at a lower price. This strategy may involve alternatives such as the employment of low wage labor, cheap imports, production abroad and government subsidies (cheap financing).

The extent to which the different alternatives can be used is of course a function of the *power structure* between the groups of providers and customers. Particularly important here are the alternatives available, i.e. what competing organizations can offer.

The above reasoning has been in terms of the 'pure strategies' and in real life we would of course expect all of them to be used in a mix. A company may thus at the same time launch e.g. marketing activities abroad, change the basic technology and start production in low-wage countries.

The discussion so far has been rather general. Having provided this general framework it now seems appropriate to turn to a particular case in order to demonstrate the possibilities of utilizing the described framework for analysis.

4. AN EMPIRICAL APPLICATION: THE NEWSPAPER CASE

The described method will be applied to the situation of daily newspapers. Their particular characteristics have been studied in another project and the main conclusions from this study (cf. Engwall [3], Ch. 9) were that *newspaper technology* is characterized by:

(a) a short production cycle

(b) a chain of tasks with different technological characteristics which have to be integrated

(c) variations in input.

It was also found that with respect to their *environment* (external providers, customers and competitors) newspapers exhibit:

(a) two types of customers (readers and advertisers) of which the advertisers are the most important ones for quality papers

(b) conservative markets, something which makes products developments difficult

(c) interest of external providers in general and owners in particular to influence the design of the product.

Finally, it was found that *newspaper personnel* are characterized by:

(a) a high degree of differentiation among as well as within personnel groups, i.e. among journalists, editorial writers, printers and administrative personnel

(b) a pronounced need for contacts between the various groups as well as between individual groups and the environment.

The different characteristics are important as we turn to analyze the situation of daily newspapers in terms of the earlier described model. It is, for instance, highly relevant to mention that the high degree of differentiation in newspaper organizations make them particularly subject to provider pressure for growth. The differentiation will increase the difficulties of sharing scarce resources. The way to avoid internal conflicts is then to try to increase the resources to be shared. For instance, if the newsroom and the advertising department or different news desks fight for space it is very likely that one tries to solve the conflict by increasing the size of the newspaper.

If the method described in the earlier section is applied to newspapers we will find that not all the substrategies listed in Fig. 2 are applicable (cf. Fig. 3).

The strategy of *increased sales to existing customers* has some limitations, since it is difficult to have an individual to read more than one copy a day of the newspaper. Some possibilities may exist of increasing the loyalty of occasional buyers or to increase the advertising input from existing advertisers. The possibilities seem, however, to be restricted.

STRATEGY ORIENTATION	AIM OF STRATEGY	WITHIN EXISTING SYSTEM	CHANGE OF SYSTEM
CUSTOMERS	To increase revenues	1. Price hikes ~~2. Increased sales to existing customers~~	1. New customers nationally ~~2. New customers abroad~~
TECHNOLOGY	To change the transformation process	Increased efficiency of present technology	Radically new technology
PROVIDERS	To decrease the demands on the organization	~~1. Reducing input prices~~ 2. Reduce input per output	~~1. Low wage labour~~ ~~2. Cheap imports~~ ~~3. Production abroad~~ 4. Government support

FIG. 3. *Possible strategies to handle organizational pressure in the newspaper industry.*

Finding new customers abroad is even more difficult in this nationally differentiated market.

With respect to the provider strategies, it should be noted that the two most important inputs of newspapers are personnel and newsprint. In terms of both these inputs one may notice difficulties in decreasing the demands on the organization.

Reducing wages as well as hiring *low wage labor* are both difficult due to the strong unions in newspaper organizations. This in turn is a result of the above mentioned characteristics of newspaper technology, i.e. that the product is produced during a short production cycle in a chain of tasks and that it is of extreme importance that the organization meets the final deadline set out by the distribution system.

Reducing newsprint prices or finding *cheap imports* are also difficult due to the cartelization of the newsprint industry.

Production abroad, finally, is not possible because of language barriers and transportation problems arising from the short production cycle. Of course, today it is possible to transfer newspaper material between countries electronically, but as far as the journalistic and the advertising material is concerned it has to be collected locally.

In relation to the remaining substrategies one may raise a number of questions. *Price hikes* is a strategy widely used with respect to prices for readers as well as advertisers. The question now is, how far will it be possible to apply this? To what extent will competition from other media be of importance?

With respect to new *customers nationally* we can again note that it has been possible to apply this strategy on the subscription market as well as the advertising market. Relevant questions to pose here are to what extent markets begin to reach saturation points, and to what extent the competition from other media will have an influence. It is also worth noting that the advertising market is very sensitive to the business cycle. One important option within the discussed substrategy may be the printing of other products than newspapers.

Another could be to identify new products, where the specific competence of the newspaper firm could be exploited.

Concerning *increases in efficiency of the present technology* it is important to stress that the printing equipment in a newspaper is generally used only during a small part of a twenty-four hour period. Possibilities exist here to increase the utilization of the printing equipment by cooperation in printing. The obstacles to such solutions are scheduling problems (i.e. the issue of which newspaper should be printed first) and fears of news leakages.

The introduction of *radically new technology* was not possible for a long time. As a matter of fact newspaper technology did not change very much from Gutenberg's time to the middle of this century. Since then, advances in computer and printing technology have provided completely new options for newspapers. This substrategy has been widely used in the last decade and will be for some time to come. The main restrictions here arise from the fact that the technological changes have repercussions for personnel. They lead to the need to solve issues such as 1. when to change and how?, 2. the distribution of benefits, 3. the demarcation between different groups, and 4. human problems (cf. Engwall [3], Ch. 7). Since these issues are difficult to solve, it is quite clear that there is a time lag between what is technically possible and what is possible in practical applications.

The substrategy to reduce *input per output* is somewhat related to technical development and is sometimes manifested through the application of the natural attrition approach, i.e. not to hire replacements when different persons retire or leave the company for other reasons. This can be expected to be the case particularly when newspapers try to cope with their problems through reduction of periodicity, i.e. by cutting down on the number of issues per week.

Another example of the same efforts involves attempts to print on thinner newsprint in order to cut down the newsprint bill. The main constraint here is that the paper must not tear too often.

Government support, finally, is a part of the existing system in many countries [9] and is an issue for debate in others (e.g. in the United Kingdom). The main question here is to what extent one can expect governments to support the newspaper industry when so many other areas require priority—or in terms of the arguments put forward in this paper—when governments are under pressure for growth to such a large extent.

In terms of the three main strategies there seem to be three important factors which have to be taken into consideration:

(1) What are the possibilities of attracting new customers?

(2) What are the possibilities of introducing still newer technology?

(3) What will the future government policy be?

It seems to be most likely that the second question will be the most important in the future for two reasons:

(a) Markets tend to reach saturation points and attracting customers away from competitors is difficult;

(b) There seem to be limits to government subsidies.

5. CONCLUDING REMARKS

The starting point of this paper was that all organizations are subject to pressures for increased compensation from their providers (personnel, financiers, governments and suppliers). This pressure adds up to a pressure for growth in revenues for the organization. Therefore it has been argued that the handling of these pressures for growth is one of the main concerns in organizational decision making. For this reason a framework for clarifying this issue has been described. It has then been applied to one particular industry, the newspaper industry. In this application the discussion has been based on earlier research findings in a project concerning organizational aspects of newspaper production. This 'theoretical' approach in the discussion may make the practioner oppose some of the arguments in the application section. This should not be considered as a significant sign of a major deficiency of the analysis, however, since what has

Vol. 6
1978

been presented in this paper is merely a device to structure the problems faced by an organization. The application section, therefore, is more of a demonstration of a methodology than a presentation of absolute answers. Indeed, if the described approach can stimulate discussions, its purpose will be achieved.

The notion that the present paper has presented 'merely a device to structure the problems faced by an organization' implies that there are many different ways to see the world and to structure the problems envisaged. That being so, it is quite natural to ask whether the approach presented here is the most relevant. In answering this question, it is appropriate to point to the economic problems faced in organizations in many countries today: markets have tended to reach saturation and discussions on growth rates concern zero-growth or even negative growth rates. In this situation, it seems important to tackle the organizational problem of pressure for growth in revenues and to focus on all possible strategies, not only the customer oriented ones.

REFERENCES

1. BARNARD C (1938) *The Functions of the Executive.* Harvard University Press, Cambridge, MA.
2. CYERT RM & MARCH JG (1963) *A Behavorial Theory of the Firm.* Prentice-Hall, Englewood Cliffs, NJ.
3. ENGWALL L (1977) *Travels in Newspaper Country.* Research Report 1/77, Department of Business Administration, University of Uppsala, Uppsala, Sweden.
4. HALL RH (1972) *Organizations.* Prentice-Hall, Englewood Cliffs, NJ.
5. HIRSCH F (1977) *Social Limits to Growth.* Routledge & Kegan Paul, London.
6. HIRSCHMAN AO (1970) *Exit, Voice and Loyalty.* Harvard University Press, Cambridge, MA.
7. MARCH JG & SIMON HA (1958) *Organizations.* John Wiley, New York.
8. MEADOWS DH (1972) *The Limits to Growth,* Universe Books, New York.
9. SMITH A (1977) *Subsidies and the Press in Europe,* Vol. XLIII (595), PEP, London.

ADDRESS FOR CORRESPONDENCE: *Lars Engwall, Esq., Department of Business Administration, Box 513, S-751 20 Uppsala, Sweden.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6, No. 6, pp. 493–500
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0493$02.00/0

Overtime Usage in a Job Shop Environment

JAMES C GOODWIN, Jr

University of Richmond, Virginia

DOUGLAS ELVERS

University of North Carolina

JACK S GOODWIN

Texas Tech University

(*Received April* 1978; *in revised form June* 1978)

Problems are encountered in a job shop which has a fixed capacity if the total work content of the jobs passing through the shop increases sufficiently. Even the use of effective priority dispatching rules and/or expediting does not adequately shorten the queues which develop if the total work content continually exceeds shop capacity. To avoid losing job orders because the orders are unduly delayed, the job shop might resort to overtime usage. This study examines the efficient and economic use of overtime to relieve the backlog problem and uses overtime as the basic criterion for evaluation of overtime usage. The study employs GPSS V programming language to simulate a hypothetical job shop. The shop is loaded to various proportions of its normal capacity and various levels of overtime are tested. Findings show that overtime should not be assigned indiscriminately but rather should be based on a shop's unique conditions of overtime cost, the priority rule being employed, and the level of capacity utilization. Marginal benefit–cost ratio curves are developed to determine whether overtime usage is economically reasonable. These curves may also be used to determine the maximum or limiting amount of overtime to use under specific shop conditions.

INTRODUCTION

MUCH of the research on job shops in the past two decades has concentrated on the dispatching aspects of the job shop. The effect of dispatching rules on different measures of shop performance criteria has been studied by numerous researchers including Baker and Dzielinski [1], Conway [3,4], Elvers [7,8], Fryer [9,10], Hottenstein [11,12], Jones [13], and LeGrand [14]. Many of these previous studies were performed under conditions where the job load throughput was held relatively constant as a fixed percentage of shop capacity (throughput load always being less than shop capacity).

One might expect that if the job shop has a fixed capacity and if the total work content of jobs passing through the shop increases sufficiently, then even with effective priority dispatching rules or with expediting, queues cannot be adequately shortened. In effect if this condition in which the total work content of jobs entering the shop continually exceeds the capacity of the shop, queues will lengthen indefinitely with jobs waiting in the queues. These long queues and the excessively long waiting periods attached to them create unwanted delays in job orders or what one might classify as a backlog in the shop. To avoid losing job orders because the orders are unduly delayed, one of a number of courses of action might be used, including overtime, extra shifts, subcontracting, and expanded capacity. Several researchers have alluded to this backlog problem including Elvers [7], Hottenstein [11], Jones [13], LeGrand [14] and Nelson [15].

It is recognized that overtime would be a short-run solution to a possible long-run problem if demand continues in excess of capacity. However, since transient overloads are com-

mon (so common in fact that they may become a way of life) and since overtime is often resorted to as at least a partial and/or temporary solution to the problem, it was decided that a systematic approach to overtime usage was appropriate. It was felt that overtime should not be used indiscriminately nor necessarily in situations where it appears to be intuitively apparent to do so. Therefore, the approach to overtime usage as set forth herein is suggested as one systematic approach to the overtime usage decision facing job shop managers.

In this study the ratio of "expected time to complete a job set" to "normal time available in the shop" is referred to as "shop capacity utilization." As shop capacity utilization is increased and the shop becomes overloaded, the backlog encountered becomes more severe if other paramaters are unchanged. Several levels of overtime (including zero overtime) are tested at different degrees of shop capacity utilization in order to relieve the backlog problem. The basic criterion for evaluation of overtime usage is tardiness.

The existence of tardiness results in some increased cost to the organization. This might be due to a number of factors such as loss of immediate demand, penalties for tardy jobs, and loss of goodwill. It is assumed throughout this study that the cost of tardiness is a linear function. The use of overtime is a possible means of decreasing tardiness and would thus benefit the organization by decreasing the aforementioned costs. However, the use of overtime also increases the organization's costs due to increased labor and overburden cost. A classic example of tradeoffs thus exists in which it is economically reasonable to increase overtime and its associated costs so long as the benefits derived from decreased tardiness exceed the costs of the overtime. The ratio of the benefits derived from reducing tardiness divided by the cost of the overtime required to effect this reduced tardiness may be viewed as a benefit–cost ratio.

The following approach attempts to examine the efficient and economic use of overtime to relieve the backlog problem. A hypothetical job shop is simulated and its performance is examined under varying conditions of priority dispatching rules, levels of overtime, and shop capacity utilization.

GENERAL DESCRIPTION OF THE STUDY

The basic model chosen represents a job shop with eight single processing facilities (machines). Each facility is different from any of the others, and each is allowed to work on only one job at a time. A set of jobs is generated randomly and released into the shop once each week. This issuance of jobs to the shop on a once per week basis is similar to the situation where the production department receives orders in the mail daily, processes these orders, and on Friday afternoon at a weekly production meeting, releases the accumulated set of jobs to the shop superintendent. Each job may require from one to four operations inclusive. Machine routing for each operation of a job is individually established. Similarly, for each operation of each job an estimated performance time is generated randomly within specified limits and is used for purposes of scheduling each job operation. After a job is loaded on a machine a subsequent actual operation time is calculated and is used to simulate a likely actual occurrence. As a job's estimated performance times are generated, a due date for the job is established. These due dates are calculated as a multiple of five times the total of the generated estimated operation times for a given job.

Since the application of the systematic approach described herein would have unique characteristics, it was not deemed necessary to give more specific information about the model used in the experiment.

The simulation models were written in the General Purpose Simulation System (GPSS) language for use on the IBM System 370/165 Computer. There were 125 separate conditions of the simulation (five priority dispatching rules × five levels of shop capacity utilization × five levels of overtime). Seven replications were made for each condition. A total of 875 simulation runs resulted in the generation of approximately 900,000 jobs with statistics being maintained on about 450,000 of these jobs.

The following five priority decision rules were chosen for the study:

(1) Earliest due date (DUEDATE)—Assign highest priority value to the jobs with the earliest due date.

Vol. 6
1978

(2) First-come, first-serve (FCFS)—Assign priority at a given machine on a first-come, first-served basis.

(3) Job slack per remaining operation (JSPO)—Relate priority value directly to the slack time per remaining number of operations with "slack" being defined as dynamic slack, i.e. time remaining to due date less the remaining process time.

(4) Random rule (RANDOM)—Assign priority in a purely random fashion.

(5) Shortest imminent processing time (SPT)—Relate priority to the processing time of the imminent operation through a direct relationship.

CRITERIA FOR EVALUATION

Students of job shop problems including Conway and Maxwell [2], Day and Hottenstein [6], Fryer [9] and LeGrand[14] have suggested and used a wide variety of criteria for evaluation. In this study, the basic criterion chosen for evaluation is tardiness.

It is generally accepted that orders which are not completed by the due date create some ill will. However, completing a job early does not necessarily create benefits. For instance, an early finish (completion prior to the due date) could result in unwanted storage costs either at the job shop or at the point of delivery. Perhaps an early finish could generate customer goodwill but it might be just as likely not to do so. In this study it is assumed that the benefit of many jobs meeting the due date offsets the benefit of a few jobs being completed extremely early. Or having a number of jobs finishing early along with several very late finishing jobs is not as desirable as having all these jobs completed near the due date. This view led to the selection of tardiness as the principal performance criterion. Tardiness statistics are then interpreted in terms of benefit-cost ratios for use in overtime usage decisions. It should be recognized that other criteria might be more meaningful depending upon the particular situation.

RESULTS OF THE EXPERIMENT

The study demonstrates how overtime can be used to reduce: (a) the percentage of jobs completed which are overdue, (b) mean job lateness, (c) mean job tardiness and (d) uncompleted (work-in-process) jobs. Not all these

TABLE 1. JOB TARDINESS STATISTICS FOR VARIOUS COMBINATIONS OF PRIORITY RULE AND OVERTIME AT 120% SHOP CAPACITY UTILIZATION

Priority rule	0%	5%	10%	15%	20%
Duedate	$x = 347$ $n = 590$	$x = 269$ $n = 569$	$x = 197$ $n = 534$	$x = 134$ $n = 463$	$x = 92$ $n = 323$
FCFS	$x = 442$ $n = 579$	$x = 350$ $n = 582$	$x = 256$ $n = 563$	$x = 180$ $n = 485$	$x = 126$ $n = 362$
JSPO	$x = 291$ $n = 588$	$x = 230$ $n = 571$	$x = 167$ $n = 539$	$x = 116$ $n = 466$	$x = 82$ $n = 301$
Random	$x = 443$ $n = 491$	$x = 378$ $n = 481$	$x = 314$ $n = 438$	$x = 252$ $n = 369$	$x = 205$ $n = 288$
SPT	$x = 437$ $n = 162$	$x = 449$ $n = 161$	$x = 458$ $n = 159$	$x = 474$ $n = 146$	$x = 393$ $n = 125$
			Overtime		

x = Mean of the mean tardiness for the seven replications (in hours).
n = Mean number of tardy jobs for the seven replications.

TABLE 2. JOB TARDINESS STATISTICS FOR VARIOUS COMBINATIONS OF PRIORITY RULE AND OVERTIME AT 110% SHOP CAPACITY UTILIZATION

Priority rule	0% Overtime	5% Overtime	10% Overtime	15% Overtime	20% Overtime
Duedate	$x = 182$ $n = 503$	$x = 126$ $n = 422$	$x = 90$ $n = 294$	$x = 64$ $n = 170$	$x = 42$ $n = 68$
FCFS	$x = 240$ $n = 524$	$x = 158$ $n = 455$	$x = 109$ $n = 332$	$x = 78$ $n = 212$	$x = 51$ $n = 121$
JSPO	$x = 158$ $n = 515$	$x = 105$ $n = 420$	$x = 75$ $n = 277$	$x = 58$ $n = 132$	$x = 42$ $n = 41$
Random	$x = 305$ $n = 414$	$x = 223$ $n = 348$	$x = 173$ $n = 267$	$x = 145$ $n = 178$	$x = 104$ $n = 125$
SPT	$x = 451$ $n = 155$	$x = 403$ $n = 150$	$x = 338$ $n = 126$	$x = 261$ $n = 89$	$x = 186$ $n = 68$

x = Mean of the mean tardiness for the seven replications (in hours).
n = Mean number of tardy jobs for the seven replications.

findings are reported herein. However, how the reduction of mean tardiness is converted into benefit–cost ratios is explained and examples of the applicability of these ratios to specific job shop conditions are presented. Although both average and marginal benefit–cost ratios were calculated and are relevant, the writers feel that the marginal ratio is especially beneficial to the job shop decision-maker.

Tables 1 to 4 show the mean tardiness and mean number of tardy jobs for various combinations of priority rule, overtime, and shop capacity utilization. Considering the DUEDATE priority rule, Table 1 indicates that at

TABLE 3. JOB TARDINESS STATISTICS FOR VARIOUS COMBINATIONS OF PRIORITY RULE AND OVERTIME AT 100% SHOP CAPACITY UTILIZATION

Priority rule	0% Overtime	5% Overtime	10% Overtime	15% Overtime	20% Overtime
Duedate	$x = 97$ $n = 321$	$x = 56$ $n = 198$	$x = 39$ $n = 73$	$x = 30$ $n = 17$	$x = 32$ $n = 2$
FCFS	$x = 120$ $n = 348$	$x = 75$ $n = 240$	$x = 50$ $n = 126$	$x = 38$ $n = 53$	$x = 30$ $n = 21$
JSPO	$x = 82$ $n = 316$	$x = 49$ $n = 184$	$x = 30$ $n = 53$	$x = 22$ $n = 5$	$x = 0$ $n = 0$
Random	$x = 179$ $n = 265$	$x = 136$ $n = 193$	$x = 96$ $n = 117$	$x = 80$ $n = 64$	$x = 60$ $n = 37$
SPT	$x = 301$ $n = 122$	$x = 234$ $n = 92$	$x = 168$ $n = 60$	$x = 120$ $n = 38$	$x = 91$ $n = 22$

x = Mean of the mean tardiness for the seven replications (in hours).
n = Mean number of tardy jobs for the seven replications.

Vol. 6
1978

TABLE 4. JOB TARDINESS STATISTICS FOR VARIOUS COMBINATIONS OF PRIORITY RULE AND OVERTIME AT 90% SHOP CAPACITY UTILIZATION

Priority rule	0%	5%	10%	15%	20%
Duedate	$x = 39$ $n = 104$	$x = 28$ $n = 21$	$x = 21$ $n = 3$	$x = 0$ $n = 0$	$x = 0$ $n = 0$
FCFS	$x = 51$ $n = 147$	$x = 38$ $n = 63$	$x = 31$ $n = 30$	$x = 24$ $n = 13$	$x = 26$ $n = 6$
JSPO	$x = 35$ $n = 68$	$x = 32$ $n = 14$	$x = 21$ $n = 2$	$x = 0$ $n = 0$	$x = 0$ $n = 0$
Random	$x = 99$ $n = 126$	$x = 77$ $n = 73$	$x = 58$ $n = 44$	$x = 47$ $n = 28$	$x = 38$ $n = 14$
SPT	$x = 171$ $n = 61$	$x = 127$ $n = 38$	$x = 99$ $n = 21$	$x = 79$ $n = 12$	$x = 54$ $n = 7$
	Overtime				

x = Mean of the mean tardiness for the seven replications (in hours).
n = Mean number of tardy jobs for the seven replications.

120% shop capacity utilization (constant in this table) and zero overtime, the mean number of jobs tardy is 590 with a mean tardiness of 347 hours. As overtime is increased in increments from zero to 20% both the mean number of tardy jobs and the mean tardiness is reduced to a level of 323 and 92 hours respectively at 20% overtime usage. This inverse relationship between tardiness and overtime generally holds true for all the priority rules and for different levels of shop capacity utilization.

A different perspective can be obtained by additional analysis of Tables 1 to 4. At the particular combination of 120% capacity utilization, zero overtime, and using the DUEDATE dispatching rule (see Table 1), 590 jobs are tardy and mean tardiness is 347 hours. Thus, viewing tardiness in the aggregate, the shop is experiencing 204,730 (590 × 347) hours of tardiness. Table 1 reveals that with 5% overtime 569 jobs are tardy with the mean tardiness reduced to 269 hours. In the aggregate the shop is now experiencing 153,669 (569 × 269) hours of tardiness. This is a net improvement (reduction) of 51,669 hours of tardiness at the 5% overtime level over the zero overtime level.[1]

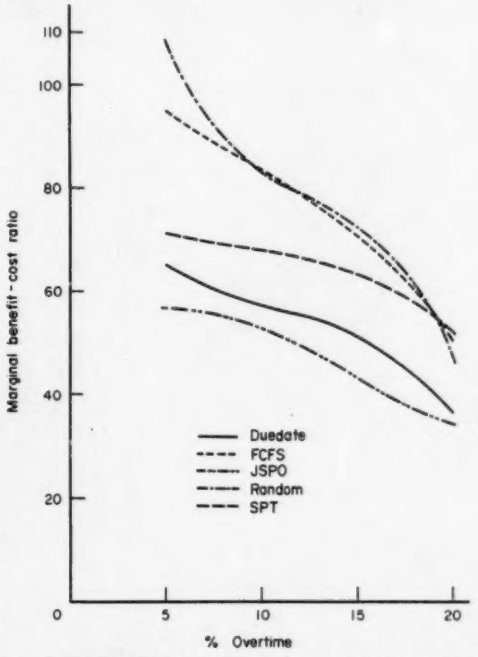


FIG. 1. *Marginal benefit–cost ratio vs. overtime at* 120% *shop capacity utilization (including WIP tardiness reduction).*

[1] As previously mentioned it is assumed that tardiness and its associated costs are linearly related. Since tardiness is viewed in the aggregate, in reality several jobs with very little tardiness could possibly be lumped together with a few jobs that are extremely tardy. In this case the slight tardiness of jobs might result in very negligible costs whereas a few jobs that are extremely tardy could be very costly to the firm. At any rate, for purposes of this study, the cost of tardiness is regarded as being linear.

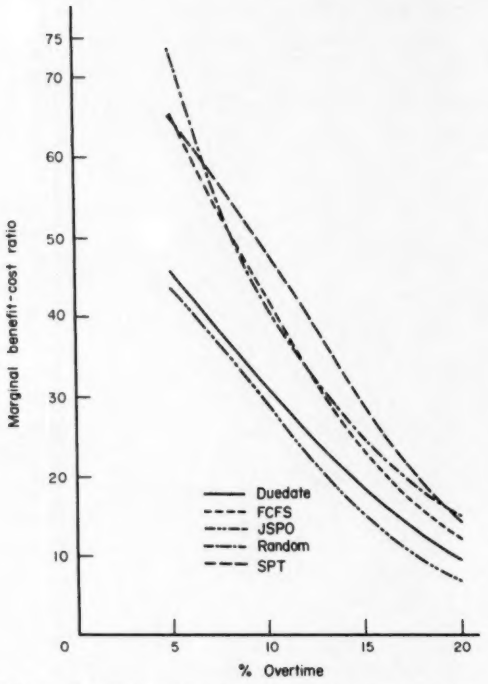


FIG. 2. *Marginal benefit–cost ratio vs. overtime at 110% shop capacity utilization (including WIP tardiness reduction).*

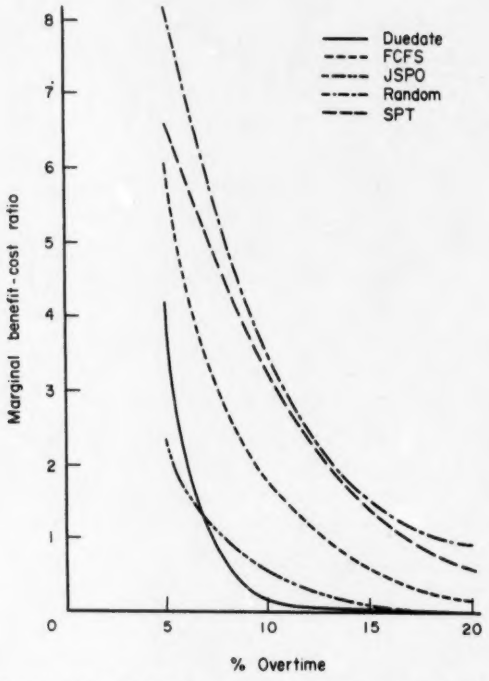


FIG. 4. *Marginal benefit–cost ratio vs. overtime at 90% shop capacity utilization.*

This improvement or benefit of 51,669 hours reduction in tardiness was not achieved without a cost. The cost to the shop for this benefit was 832 overtime hours. The benefit–cost ratio in terms of hours saved (reduction of tardiness) to hours expended (increase in overtime) is 51,669 ÷ 832 or 62.1. Or one might say that for each hour of overtime invested on the average, the shop experiences 62.1 fewer hours of tardiness. These benefit-cost ratios may be noted in Figures 1 to 4.

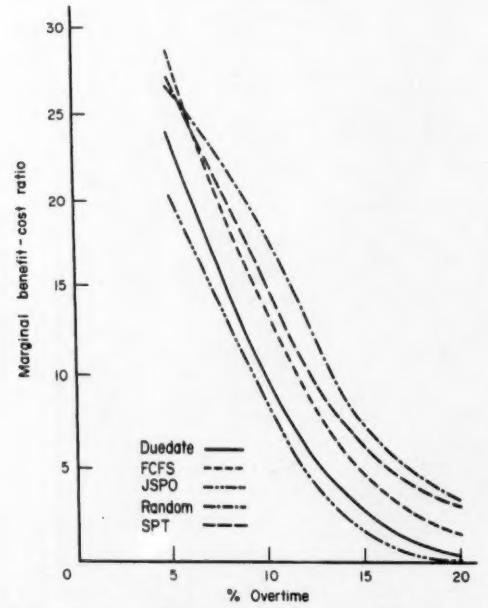


FIG. 3. *Marginal benefit–cost ratio vs. overtime at 100% shop capacity utilization (including WIP tardiness reduction).*

It might be concluded then that overtime usage in this particular situation is a very favorable course of action to follow since the returns appear to be so much greater than the expenditures (in terms of hours). However, one should not reach this conclusion *only* on the basis of the benefit–cost ratio expressed in terms of hours. The key to the decision of overtime usage lies in the *cost* of the hours involved here because if the cost of one hour of overtime is not equal to the benefit of reducing tardiness by one hour then further interpretation of the benefit–cost ratio is necessary before a decision to use overtime can be made. One must consider the cost of overtime usage and the benefit of reducing tardiness in terms of *dollar amounts* before an economic decision can be reached. Naturally the benefit to be derived should exceed the cost before it becomes clear that overtime should be used.

Numerous applications of the benefit-cost ratios of Figs 1 to 4 exist. One such application follows. Suppose the shop in our simulation applied 5% overtime for one week. This will

Vol. 6
1978

cost the shop 16 hours of overtime. With the aforementioned 62.1 benefit cost ratio, the shop would benefit by reducing tardiness by 993.6 hours (16 × 62.1). If overtime costs $5 per hour then the total cost of 16 hours of overtime would be $80. The decision as to whether this use of overtime is advantageous to the shop depends on the benefit to the shop of 993.6 hours reduction of tardiness. Is this reduction in tardiness worth more than $80 to the shop? Obviously, if the answer is yes, then the use of overtime is beneficial.

When the dollar value of benefits and costs are known or can be estimated, this ratio of dollar value of benefit/hour divided by dollar value of cost/hour is combined with its inverse benefit–cost ratio (in terms of hours) resulting in a point of indifference equal to one. Figures 1 to 4 can then be used to locate the upper economic limit of overtime usage. As in all marginal economic analysis, this is the point of overtime usage which results in *optimum* benefit to the firm since marginal costs are equal to marginal benefits.

A note of precaution is appropriate here. In the evaluation of overtime decisions, care must be given as to whether overtime costs are constant over all ranges. In other words, to move from no overtime to 5% overtime might cost $6 per hour (assuming 1-1/2 times normal cost of $4/hour), whereas moving from 15% overtime to 20% overtime might cost $8 per hour (2 times normal cost of $4/hour). This is especially important in transforming the benefit–cost ratios in Figs 1 to 4 from hourly to dollar figures. Without taking this possible change in overtime cost from one level to another into account, error in interpretation might result.

SUMMARY

An important point underscored by this research is that overtime should not be used indiscriminately or in situations where it appears to be intuitively apparent to do so. An evaluation of the priority dispatching rule being applied in the shop, the level of shop capacity utilization, and the cost of overtime in the shop is necessary to interpret whether overtime is an economical alternative to relieving the backlog in the job shop.

Overall it would seem that overtime is a very effective and economically feasible action to employ under some backlog conditions. Using this specific model as an example, many of the benefit–cost ratios are in the 50 to 1 and up to 100 to 1 range. Cost of overtime would have to be very high and/or the benefits derived from reducing tardiness would have to be very small in order for a shop under sufficient backlog to ignore the use of overtime at least on a temporary basis.

A practitioner could use this approach to determine the specific benefit–cost ratio for his particular conditions of priority rule, level of shop capacity utilization and overtime level desired. Then applying the costs of overtime and benefits of reducing tardiness to this ratio (as previously explained), it can be determined whether use of overtime is an economically reasonable choice. This is the basic contribution of this research to the study of job shops.

In addition, marginal benefit–cost ratio curves similar to those presented in Figs 1 through 4 may be used to determine the maximum or limiting amount of overtime to use under specific shop conditions of priority rule and shop capacity utilization. This limit of overtime usage is also the optimum economic amount of overtime to use under the specific shop conditions.

In conclusion, it should be noted that the use of overtime and its related benefits should not be thought of in terms of a one to one correspondence. Rather there exists a multiplicative effect as regards overtime benefits. Not only does the job presently being attended to receive benefits, but various other jobs in the queue and possibly elsewhere in the shop receive residual effects of overtime usage. The simulation approach to overtime usage suggested herein allows one to systematically analyze the benefits to be derived and their inherent costs.

REFERENCES

1. Baker GT & Dzielinski BP (1960) Simulation of a simplified job shop. *Mgmt Sci.* **6**, 311–323.
2. Beaird SD (1968) A practical approach for job shop planning and control (part 2). *Automation, Cleveland*, **15**, 92–98.
3. Conway RW (1965) Priority dispatching and job lateness in a job shop. *J. Ind. Engng*, **16**, 228–237.
4. Conway RW (1965) Priority dispatching and work-in-process inventory in a job shop. *J. Ind. Engng*, **16**, 123–130.
5. Conway RW & Maxwell WL (1962) Network dis-

patching by the shortest-operation discipline. *Ops Res.* **10,** 51–73.
6. DAY EJ & HOTTENSTEIN MP (1970) Review of sequencing research. *Nav. Res. Logist. Q.* **507,** 11–39.
7. ELVERS DA (1973) Job shop dispatching rules using various delivery date setting criteria. *Prod. Invent. Mgmt,* **14,** 62–70.
8. ELVERS DA (1974) The sensitivity of the relative effectiveness in job shop dispatching rules with respect to various arrival distributions. *AIIE Trans.* **6,** 41–49.
9. FRYER JS (1971) A simulation study of a hypothetical job shop with machine and labor constraints and two levels of labor control. Unpublished doctoral dissertation, Indiana University, USA.
10. FRYER JS (1974) Labor flexibility in multiechelon dual-constraint job shops. *Mgmt Sci.* **20,** 1073–1080.
11. HOTTENSTEIN MP (1969) A simulation study of expediting in a job shop. *Prod. Inv. Mgmt* **10,** 1–11.
12. HOTTENSTEIN MP (1970) Expediting in job-order-control systems: a simulation study. *AIIE Trans.* **2,** 46–54.
13. JONES CH (1973) An economic evaluation of job shop dispatching rules *Mgmt Sci.* **20,** 293–307.
14. LEGRAND E (1963) The development of a factory simulation system using actual operating data. *Mgmt Technol.* **3,** 1–19.
15. NELSON RT (1967) Labor and machine limited production systems. *Mgmt Sci.* **13,** 648–671.

ADDRESS FOR CORRESPONDENCE: *Professor James C Goodwin, School of Business Administration, University of Richmond, Virginia* 23173, *USA.*

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 501–506
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0501$02.00/0

Goals and Realities in a Distribution System: The Case of Pharmaceuticals in Norway

KJELL GRØNHAUG

Norwegian School of Economics and Business Administration, Bergen, Norway

(*Received June* 1978)

As several writers have indicated, significant deviations from the competitive model are found in pharmaceutical markets. Aside from barriers related to scale economies, technological factors and demand conditions, a variety of legal restraints are placed on production, distribution and consumption of such products. A description of the distribution of pharmaceuticals in one country (Norway), has been undertaken, and the impact of specific restrictions on the conduct of various factors and some aspects of performance in this distribution system are discussed and analyzed.

INTRODUCTION

THE MAIN purpose of this paper is to focus on the impact of various government regulations on the conduct and performance of companies in the distribution of pharmaceuticals in Norway.

The final goal of all productive and distributive activities, most marketers believe, is to satisfy the ultimate consumer. Perfect competition—under its heroic assumptions is believed to lead to optimal allocation of resources and to maximize social welfare [5]. Several writers [1, 4, 13, 15, 27], however, who have studied the pharmaceutical industry, have noted considerable deviations from the competitive model in this industry. These deviations relate to the number and size of the actors in the industry, the heterogeneity of products, and lack of product information for consumers in addition to numerous regulations. All these factors affect both the structure and conduct of companies in the industry. It should be noted, however, that actual regulations and structure—and thus conduct and performance—vary considerably among countries. The paper is organized in the following way:

- The Norwegian pharmaceutical distribution system will be described.
- The stated goals and premises in this system will be identified, and
- Criteria, designating performance standards, and empirical evidences related to that performance will be presented and analyzed.

THE DISTRIBUTION SYSTEM

Figure 1 gives a simplified representation of the actual distribution system:[1]

Supply

The production of pharmaceutical products takes place in industrial firms and pharmacies, with the former accounting for 81% of production and the latter 19% (1975). The production in pharmacies is restricted to products listed in a published pharmacopia. Only about one fourth of the industrial supplied pharmaceuticals has domestic origin. The domestic producers are enjoying a modest, but growing export trade.

Distribution

All the domestic wholesale trade as well as imports and exports of pharmaceuticals are controlled by NMD (The Norwegian Medical Depot). Retail sales are restricted to pharmacies and authorized medical outlets.

The number of pharmacies has been almost

[1] Figure 1 reflects only the main elements in this distribution system. Several associations and activities are left out. For a more detailed description of this system cf. [10, 22, 28].

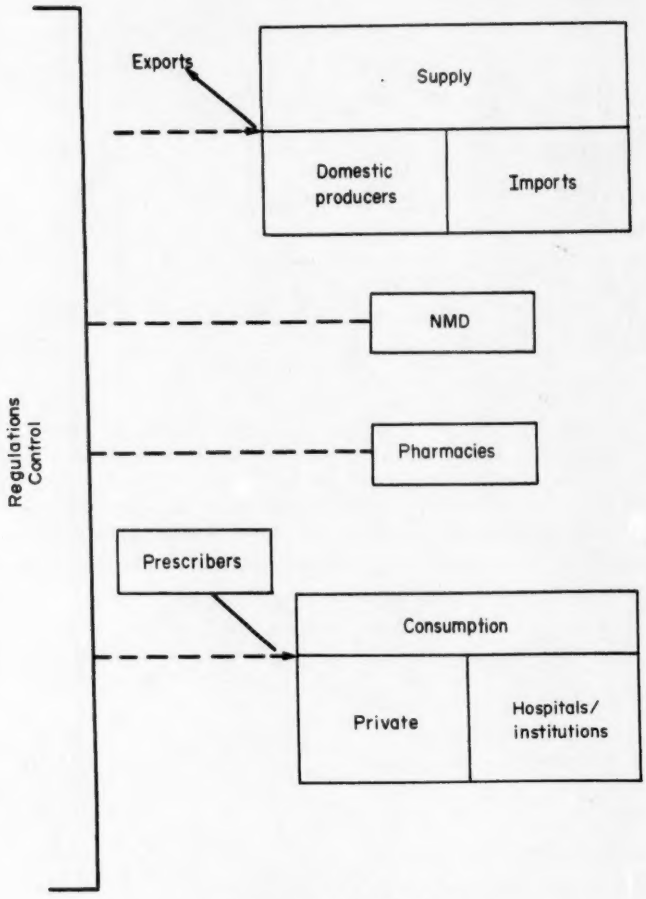


FIG. 1. *The distribution system.*

constant during the last decade (approximately 275–285) as has the medical outlets (approximately 1,000–1,100).

Consumption

Use of pharmaceuticals may be divided according to two distinctions. The first relates to whether they have to be prescribed by a physician or not. The second relates to the place of consumption, i.e. either at home, in hospital or other institutions.

Control and regulations

All of the elements in the distribution system as depicted above, are subject to a variety of legal restrictions.[2] Both domestic and foreign producers are restricted in several ways.

- Foreign producers have to be represented by a domestic agent in order to sell their products in this market.
- Both the establishment and the operation of pharmacies (and authorized outlets) are constrained.
- Introduction of new products, prices and price changes, and advertising and other forms of promotion, are subject to control.
- The day-to-day-operations also are watched by public controllers.
- Consumption is regulated by requiring a prescription for many pharmaceuticals. In addition the transfer of a prescribed drug to other persons is prohibited.

[2] No attempt will be made here to give a detailed description of the various laws regulating this distribution system. Attention should, however, be paid to Law of drugs affecting both production, distribution and consumption of pharmaceuticals [14].

GOALS AND PREMISES

The prime goal in this distribution system seeks to *serve* the ultimate consumer. This emphasis on insuring good *health* along with political goals related to *equality* in distribution of income, have heavily influenced regulatory activities regarding the distribution of pharmaceuticals.

With respect to this industry, the public view is as follows: Good health is a desired state. There is a high probability that most individuals will be sick, and thus have a need for pharmaceuticals. That is, pharmaceuticals are needed to reestablish the desired state (good health) and bolster economic wellbeing. Thus pharmaceuticals are regarded as important and thus become subject to public scrutiny. Furthermore, various assumptions regarding the functioning of the system are subsumed under the label "*lack of information*". For example, it is assumed that the individuals are *unable* to act according to their own real interests, i.e. determine their real need for pharmaceuticals. Furthermore (as will be discussed below) the informational capacity of the prescribers is also assumed to be limited.

Innovations, i.e. new and better pharmaceuticals are also wanted. For strategic and economic reasons a high fraction of the consumption is wanted to be of domestic origin. From the stated goals (good health, equality) and premises (lack of information), and preference for domestic production, several subgoals can be deduced:

A. Effective, safe and better products.

B. Accessibility and guaranteed supply.

C. Low and fair prices.

D. Correct use.

These goals may furthermore be related to more specific goals and actions:

A. *Products*:

1. Restricted entry.
2. Extensive control over production.
3. Extensive control regarding the efficacy and safety of products before they are allowed to be sold.
4. The desire for new (and better) products.

B. *Accessibility*:

1. The pharmacies have to carry a certain assortment (assortment standards).
2. Emphasis is on domestic production (to ensure supply).

C. *Prices*:

1. Price controls on new products (producers).
2. Control over price changes (producers).
3. Controls over margins (distributors).

D. *Correct use*:

1. Prescription is required for many products.
2. Control over prescribers and pharmacies.
3. Rules regulating consumption (which prohibit circulating prescribed products to another person).
4. Control on the number of products marketed, in order to facilitate the choice situation for the prescribers.
5. Controls on market communications to prescribers and users (i.e. advertisements, brochures, etc., have to be approved by the official control institution before use).
6. Salesmen visiting prescribers have to pass a specific examination.
7. Information to consumers about the use of pharmaceuticals.

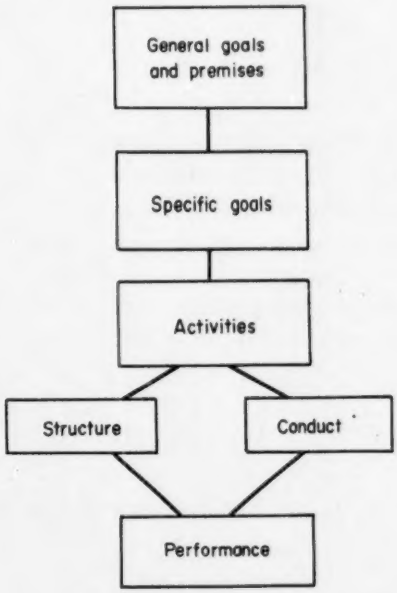


FIG. 2. *Goals, activities and performance.*

In other words, the general goals and premises may be seen as affecting more/specific goals and activities. A closer look at these activities reveals that they do affect both the structure and the conduct of actors in order to influence the performance of this distribution system.

PERFORMANCE

Performance is often thought of in terms of either prices, or related questions like: "Do the pharmaceuticals cost too much?"; and "Why do prices of pharmaceuticals vary across countries?" [4, 13]. Performance, however, is a *multidimensional* concept [26, p. 37–38]. It seems reasonable to link performance to stated goals for the distribution system.

1. *Safe products*

In this controlled distribution system, products with serious side effects have been marketed. However, when considering serious cases (cf. Thalidomide), cross-national comparisons indicate this to be less pronounced than in most other Western countries.

2. *Effective products*

Does this system promote effective products? Highly effective products occasionally will have negative side effects associated with them. And with the emphasis on safe products, many effective products (with negative side effects) will probably *not* be marketed. In statistical term, this system favors type II errors.

3. *New products* (*innovations*)

Given the reasons out-lined above, the system probably will not favor innovations, neither in terms of creating new products, nor in early adoption of such products [9].

4. *Access*

If equality in access exists then the consumption of pharmaceuticals across geographical areas should be approximately the same, assuming the same relative distribution of various illnesses. However, when looking at consumption data for various regions, differences emerge that cannot be accounted for by age, and variations in propensity to be attacked by various illnesses.[3] The overall result is that the per capita consumption correlates positively with population density. Similar results are reported for selected products by Bergman [2].

5. *Domestic production*

To guarantee supply, the government wants production to be mostly of domestic origin. However, the proportion of domestic production has been declining for the last decades. From 1963 to 1975, the domestic proportion of pharmaceuticals consumed fell from approximately 44% to 25%. The number of domestic producers also has declined, and their production has been concentrated on certain specialities. In other words, the more important product groups are supplied by foreign producers, thus making the supply situation more vulnerable.

6. *Prices*

As noted earlier, pricing of pharmaceuticals generally is controversial, and such is the case in this distribution system.[4] In comparing the drug price index with the consumer price index, the indices showed 120 and 165 respectively in 1975 (1968 = 100). On the other hand, the consumer's money outlay has increased more [23]. This is partly because of an increase in consumption. However, the consumption of important pharmaceuticals (NMD 1977) for the period 1972-1977, shows only a modest increase in consumption. Furthermore, the price for several of the most important product groups (measured in terms of sales volume), have increased significantly more than indicated by the price index for pharmaceuticals. The apparent inconsistency may be explained in the following way: Because of a short product life-cycle, the index only maps a modest fraction of the actual consumption, and thus inadequately represents the actual consumption. As noted above, the prices have to be approved before a new or modified product is placed on the market. This, however, is a very difficult task—in particular when confronted with a multinational producer. Because of the difficulties in assessing costs and pro-

[3] Even though statement is based on survey data using very aggregated consumption measures [23], the data leave almost no doubt that such differences really exist.

[4] For an overview of the basic issues in this debate cf. [6, 7, 16, 28], and the newly created newsletter issued by the association representing foreign firms (No-Re-Farm).

fits, it is difficult to say how effective the price-control is in fact [11].

7. *Correct use*

Several questions relating to the use of pharmaceuticals have been examined in some studies (for overview cf. [8]). The overall impression is that incorrect use (i.e. overconsumption, underconsumption, use of pharmaceuticals prescribed for other persons etc.) occurs to some extent. Information obtained through surveys also reveals that the consumer's level of information (related to labeling, use, stocking etc.) is modest.

8. *Limited number of products*

Compared to other countries the total number of products is very low. The number of registered pharmaceutical specialities was 1,845 in 1975, representing about 750 drugs.[5] This compares with approximately 20,000 in the US. Furthermore, since 1970, the number of registered specialities has been declining.[6]

DISCUSSION

1. The preceding section revealed both multiple goals and premises, and multiple criteria of performance. Furthermore, some of the goals may be perceived as *conflicting*, i.e. the desire for safety may conflict with the desires for effectiveness and innovation.

2. Another point is the *variability* with regard to degree of specification across the various subgoals, which probably is related to the level of control exerted. Exerting influence over the number of products on the market is definitely easier (more operational) than exerting influence over the rate of innovation. Furthermore, emphasis seems to be on goals and criteria which are easy to control, which, is in accordance with the well-known "Gresham's Law" of planning [17, p. 186–87], but which promotes deviances from the less specific goals.

3. An important aspect is that actions undertaken to pursue specific goals may produce negative *side-effects* contradicting other goals. To be more specific:

- Restrictions on number of products may have negative effects on the introduction of new products.

- Furthermore, by setting limits on the number of products in each product category, the probability of acceptance will be positively related to *how fast* a new product is created. New product development is positively related to resources [12]. In other words, smaller firms, low in resources seem to be the most vulnerable to this restriction. Since domestic firms are lower in resources than foreign companies, their market share is decreasing.

- Since prices are controlled, it is easier to control price-changes for old products than to control and/or judge the fairness of prices for new products. Big firms with more resources introduce more new products than smaller firms, and thus are less affected by the price control. Consequently, restriction on the number of products and price controls are likely to produce substantial negative effects on firms with less resources.[7] The decline in market share for domestic producers, the reduction in the number of firms, and a limited number of product may be taken as "prima facie" evidence partially supporting this assumption.

- A large portion of the pharmacies are not making adequate profits. Pharmacies making low profits or losing money are subsidized. The restrictions on the pharmacies' assortment policies may contribute negatively to the economic results, which in turn, may lead to higher prices.

4. The underlying assumption behind the attempts to keep down the number of products is that this will help the prescribers to make easier and better choices (prescriptions). But will such a restriction, i.e. 1,800 compared to 4,000 products actually affect the prescribers' choice? Findings from cognitive psychology indicate that the actual number of alternatives

[5] One drug may occur in different forms and with different potencies. Each form or potency is counted as a separate pharmaceutical speciality.

[6] Based on data from the control institution (Statens Legemiddelkontroll).

[7] For a detailed discussion related to policy outcomes cf. [3].

taken into account is limited (cf. "The Magical Number Seven Plus or Minus Two" [18]). This seems to be the case in this market. Trost [30] found that the physicians used only a small fraction of the accessible products—a result which is in accordance with the findings regarding size of "evoked set" in consumer behavior.

5. Negative discrepancies appear with regard to the wanted and the actual level of information and use behavior. Many questions about what is consumed by whom still remain unanswered. Furthermore, few attempts to specify relevant criteria and data have been made.

The *spatial* inequality in consumption should also be noted. There is some evidence that the region (rural) overlaps with age (high) and income (low) which both are related to consumption of pharmaceuticals. Thus a series of factors may attribute to inequality.

6. When taking a closer look at the regulatory activities affecting this distribution system, the primary impresssions are:

- the various goals are only *superficially* discussed and not operationally defined (and thus it is impossible to achieve a consensus as to what should be accomplished);
- there is a complete *absence* of effort in tracing relevant performance criteria;
- no real discussion nor analysis of *data* are used to evaluate performance.

The lack of operational (well defined) goals leads to inability to determine relevance, performance criteria and hence, there is no evaluation. Furthermore, the lack of relevant feedback makes it difficult to interpret the past, hampers adequate learning, and thus, contributes to a stable and biased model of this system.[8] Because of the ambiguity involved, emphasis will probably be placed on activities consistent with the views of the dominant political coalition, and less on the actual effects of the various regulatory activities.[9]

[8] For a detailed discussion of such a perspective [17].
[9] Recently this has been recognised as a problem [24].

REFERENCES

1. Arrow K (1963) Uncertainty and the welfare economics of medical care. *The Am. Rev.* **L111**(5), 941–973.
2. Bergman U *et al.* (1975) The measurements of drug consumption. *Europ. Jl Clin. Pharmacol.* **8**, 83–88.
3. Cook TJ & Scioli FP (1972) A research strategy for analyzing the impacts of public policy. *Admin. Sci. Q.* **17**(3), 328–339.
4. Cooper MH (1966) *Prices and Profits in the Pharmaceutical Industry*. Paramon Press, New York.
5. Ferguson CE (1969) *Microeconomic Theory*. Richard D Irwin, Homewood, IL.
6. Forbrukerrapporten (1976) Medisiner for millioner. *Forbrukerrapporten* No. **7**, pp. 24–31.
7. Forbrukerrrrapporten (1977) Store prisforskjeller på medisiner i Europa. *Forbrukerrapporten* No. **7**, p. 29.
8. Gugelli P (1976) Legemiddelforsømmelse. *Tidsskrift for Den norske lægeforening*, No. **9–10**, 545–547.
9. Grabowski HC & Vernon JM (1977) Innovation and invention. Consumer protection regulation in ethical drugs. *Am. Econ. Rev.* Paper and Proceedings **67**(1), 359–364.
10. Grønhaug K (1967) Distribusjon av og informasjon om legemidler (unpublished report) Stvgr.
11. Halse M (1975) Priskontroll på legemidler i Norge. *Nordisk Medicin* **90**(3), 92–93.
12. Kamien MI & Schwartz NL (1975) Market structure and innovation: a survey. *Jl of Econ. Lit.* **XIII**(1), 1–37.
13. Kefauver E (1966) *In a Few Hands*. Penguin Books, Hammondsworth, UK.
14. *Law of drugs* (Lov om legemidler, gifter m.v. av 20. juni 1964 med endringer ved siste lov av 19. juni 1969).
15. Lilja J (1975) *Läkarens läkemedelsval ur samhällets synsvinkel*. Bonniers, Stockholm.
16. Lillebø A (1976) Statlig kontroll fra innkjøp til salg. *Økonomisk Rapport* No.17, 8–15.
17. March JG & Simon HA (1958) *Organizations*. John Wiley, New York.
18. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* Vol. 63, 81–97.
19. NMD (Annual reports 1970–1976).
20. NOMI (Annual report 1971–1976).
21. No-Re-Farm (1977) *Legemidler, industri og samfunn*, No-Re-Farm, Oslo.
22. No-Re-Farm NB! (newsletter), Vol. 1, No. 1–5.
23. NOS (1975) *Forbrukerundersokelse* 1973, A705, Statistisk Sentralbyrå, Oslo.
24. NOU (1973) *Administrative og okonomiske konsekvenser av lover m.v.* NOU 1973:52, Oslo Universitets forlaget.
25. March & Olsen JP (1976) *Ambiguity and Choice in Organizations*. Universitets forlaget, Oslo.
26. Scherer FM (1971) *Industrial Market Structure and Economic Performance*. Rand McNally, Chicago.
27. Slater B (1977) *Competition and Marketing Strategies in the Pharmaceutical Industry*. Croom Helm, London.
28. Sosialdepartmentet (1972) *Legemiddelkontrollens fremtidige organisasjon* (official report), Oslo.
29. Statens Legemiddelkontroll (Annual reports 1974–1975).
30. Trost J (1961) *Informationsspridning till läkarkåren och inom den*. Sosiologiska institutionen, Apothekernes Laboratorium, Uppsala, A.S.

ADDRESS FOR CORRESPONDENCE: *Kjell Gronhaug Esq., Associate Professor, Norwegian School of Economics and Business Administration, Institute of Marketing Economics, Helleveien 30, N-5000 Bergen, Norway.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 507–513
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0507$02.00/0

Input-Output Analysis and Management Accounting: An Application Concerned with Scrap Recycling

B HARDING[1]

JI ROUND[1]

University of Warwick, UK

(*Received June* 1978)

This paper outlines the use of an input–output model in the management accounting process for a vertically integrated group of companies. In particular, it emphasises the way in which the 'through-group' cost structures may well differ from those of individual members of the group taken in isolation. Some modifications to the basic model are proposed so as to accommodate the recycling of scrap in the context of an engineering group. This provides a means of instantly determining first, the amount of externally supplied scrap, and secondly, an approach to allocating the benefits of scrap recycling to member companies when viewed from a group-wide standpoint.

1. INTRODUCTION

MANY vertically integrated groups of companies have, in a variety of ways, relatively complex structures. Some of the inherent difficulties of management accounting can be facilitated, if not even resolved, by formal analysis of the underlying structure. The very nature of vertical integration implies a good deal of interdependence between the individual companies. Existing techniques, such as input–output analysis, have been shown to provide useful insights into questions of interdependence at the economy-wide level but have received relatively little attention at the corporate level. This paper outlines some applications of this technique within the latter context and pays particular attention to the problem of scrap recycling.

The paper is in two parts. In the first section some features of a simplified form of a micro input–output model will be described and following this, in the second section, some particular modifications will be discussed which more adequately treat the problem of scrap recycling within an engineering context. More generally, however, it is hoped that this example will serve to illustrate how a relatively simple technique can be used quite effectively in the corporate planning and accounting process.

2. AN INPUT–OUTPUT MODEL FOR MANAGEMENT ACCOUNTING

In this section, a simplified version of a micro input–output model will be discussed, together with an illustration of its use in determining company cost structure in a wider 'through-group' context. Schematically, the input–output accounts for a system of n product groups can be represented as shown in Table 1. For present purposes, a product group is defined to be a part of a company, or group of companies, producing a single product. The entries in Table 1 show the structure of sales

[1] The authors are, respectively, a Corporate Planner, BL Cars, and lecturer in Economics, University of Warwick, UK. This paper is based on part of an MSc. Thesis in Business Studies submitted by B Harding to the University of Warwick (1977) prior to his appointment at BL Cars. The authors accept sole responsibility for its contents.

TABLE 1. INPUT-OUTPUT STRUCTURE OF ACCOUNTS FOR PRODUCT GROUPS

	Product groups	Non-group customers	Total sales
Product groups	X	Y	$\mathbf{g}$
Non-group purchases	W	—	$\mathbf{w}$
Total inputs	$\mathbf{g}$	$\mathbf{y}'$	ϕ

and purchases for each of the product groups. Thus, the product group rows show sales to other products within the group of companies concerned (X) and to customers outside the group (Y). Likewise, the product group columns show purchases of goods and services produced by other members of the group (X) together with other items of costs (W) detailed below. The precise definition of the matrices shown in the cell entries is as follows.

X is an $n \times n$ matrix of element $\{x_{ij}\}$ showing all the direct flows between the n product groups.

Y is an $n \times h$ matrix of element $\{y_{ij}\}$ containing details of all sales by product groups to non-group customers analysed by h main end markets. One such breakdown with which the authors are familiar is as follows:

(a) Passenger Cars—U.K.
(b) Commercial Vehicles—U.K.
(c) Tractors and Agricultural Machinery—U.K.
(d) Building and Construction—U.K.
(e) General Engineering—U.K.
(f) Passenger Cars—Export.
(g) Commercial Vehicles—Export.
(h) Tractors and Agricultural Machinery—Export.
(i) General Engineering[2]—Export.

W is a $k \times n$ matrix of element $\{w_{ij}\}$ containing details of the cost structure by product group, excluding purchases of goods and services produced by other members of the organisation. One possible set of classifications for such a model would be:

(a) Purchase of scrap.
(b) Purchase of steel billet.
(c) Purchase of other steel.
(d) Purchase of other material.
(e) Direct labour.
(f) Variable overhead.
(g) Gross margin.

Clearly, items (a)–(d) represent 'outside' purchases of materials, supplementing those produced within the group.

The total value of sales for the product groups are represented as elements of the vector $\mathbf{g}$ which must be equal to the total costs (inclusive of gross margins) shown in Table 1 as the row vector $\mathbf{g}'$. The total group cost structure is given by $\mathbf{w}$, and the total purchases of group products by the various non-group customers is shown as the row vector $\mathbf{y}'$. ϕ is simply an element portraying the sum of total sales or total inputs.

These matrices constitute the principal input data to the system but a number of peripheral (yet equally important) data sets can be held in store to be used in conjunction with them. These could include a row vector of working capital requirements by product group, forecasts of sales growth in constant price terms by market and by product group and, in certain cases, information on available annual capacity over a five year time horizon.

Coefficient forms for the matrices shown in Table 1 can be defined as follows.

$$A = \{a_{ij}\} = \{x_{ij}/g_j\}, \forall_{ij} \quad (1)$$
$$B = \{b_{ij}\} = \{y_{ij}/y_j\} \quad (2)$$
$$C = \{c_{ij}\} = \{w_{ij}/g_j\} \quad (3)$$

Note that A sets out the within-group purchases per unit output of the product groups, while C shows the outside of group purchases (including wage costs and margins) as a proportion of the same output. Matrix B simply shows the per unit demands for group produced commodities in each of the separately designated markets. The columns of B therefore sum to unity.

The input–output flow balance can be

[2] The general engineering headings are, in fact, residual classifications into which all sales not otherwise classified are placed.

Vol. 6 1978

TABLE 2. MATRIX MULTIPLIERS IN THE MODEL

	Product groups (n)	Non-group customers (h)
Product groups (n)	$I - A)^{-1}$	$(I - A)^{-1} B$
Non-group purchases (k)	$C(I - A)^{-1}$	$C(I - A)^{-1} B$

expressed as

$$\mathbf{g} \equiv \mathbf{Ag} + \mathbf{By}$$

or, solving for **g**,

$$\mathbf{g} \equiv (\mathbf{I} - \mathbf{A})^{-1} \mathbf{By} \quad (4)$$

Now under strict assumptions about the constancy of the matrices A and B, the output levels of the product groups, **g**, necessary to meet a new vector of demands in the main end markets, $\tilde{\mathbf{y}}$, is given by

$$\mathbf{g} = (I - A)^{-1} B\tilde{\mathbf{y}} \quad (5)$$

Four matrices which are variants of the multiplier matrix shown in equation (5) are expressed in Table 2.

These matrices have row and column dimensions according to the block in which they are contained and may be interpreted as follows.

(i) The columns of $(I - A)^{-1}$ show the direct and indirect (multiplier) effects of a unit increase in the output of product groups.

(ii) Equation (5) has already indicated that the columns of $(I - A)^{-1} B$ give details of the total output requirement from each product group needed to service a unit demand from each main end market. It should be noted that the assumption of a constant matrix B is particularly strong, although it enables a measurement of broad orders of effect very rapidly.

(iii) The columns of $C(I - A)^{-1}$ gives an estimate of 'through-group' cost structure by product group. They show the overall (group) cost structures arising out of unit outputs of each product group, and consequently provide central management with a 'group' view of profitability.

(iv) The columns of $C(I - A)^{-1} B$ give details of the 'through-group' cost structure by market. The information indicates (at group level) the relative profitability, labour intensity, etc., of each main end market. With changing product mix over time (i.e. changing B) this multiplier matrix would quickly provide new estimates of these group-wide indicators.

The peripheral information, referred to earlier, can be used in conjunction with one or other of these multiplier matrices to generate estimates of the effects of various changes on capacity utilisation, profit generated, working capital requirements, and many other profiles of special interest to accountants and corporate planners.

3. SOME MODIFICATIONS TO ACCOMMODATE THE RECYCLING OF SCRAP

In principle, the matrix X, together with the matrix A derived from it, includes all of the flows of goods and services between the various product groups. However, the sale of scrap from the intermediate and final production processes to the primary production processes, although apparently illustrating a good example of technical linkage in a group context, requires a special treatment in the model. The standard input-output format illustrated in section 2 is inappropriate in this particular instance. In this section a modification to the model will be illustrated by means of a simplified hypothetical example.

Suppose that a group consists of four companies each concerned with the manufacture of a single homogeneous product. These products will be referred to as billet, rod, wire and fasteners, and the pattern of direct sales both within the group and externally are as shown in Table 3. This is the standard micro input–output format. The entries in the first column which intersect with rows Q, R and S, represent flows of scrap into product group P. An inversion of $(I - A)$ based on this table would produce unsatisfactory results on at

TABLE 3. STANDARD MODEL FORMAT

Sales from \ Sales to	Product group P	Q	R	S	External sales	Total sales
P (Billet)	—	200	200	—	100	500
Q (Rod)	10	—	200	300	300	810
R (Wire)	20	—	—	300	500	820
S (Fasteners)	20	—	—	—	1000	1020
Costs (external)						
Scrap	100	—	—	—		
Other Material	—	200	100	—		
Labour	200	200	200	250		
Margin	150	210	120	170		
Total	500	810	820	1020		

least two grounds. First, the formulation would imply that an increase in billet production always generates proportional increases in inputs from the rod (Q), wire (R) and fastener (S) processes. But as noted above, these are simply requirements of scrap by the billet process. Scrap will not be supplied by processes Q, R and S if there is no prior demand for their principal products. It is also unrealistic to assume that scrap is supplied in fixed proportional amounts by the three processes. Secondly, the group financial effects of a rise in billet sales would be inaccurately apportioned between companies. To enable comparisons with a revised formulation, the through-group cost structures that would be obtained without regard to these deficiencies are shown in Table 4. The entries comprise the elements of matrix $C(I - A)^{-1}$.

The problem of finding an appropriate treatment of scrap recycling falls within the general secondary product problem in input–output analysis [1]. One possible solution is to introduce a fictitious process into which all scrap is sold and from which the sum of all scrap sales to the billet process would be made. The cost structure of this fictitious process would contain a negative gross margin which would remove the tendency of the overstatement of 'through-group' gross margins for the other product groups. However, the method would continue to produce distorted estimates of the demand for scrap from within-group sources, and hence of the billet process's scrap purchase requirement from non-group sources.

If the model is to cope satisfactorily with different rates of expansion in sales to main end markets (external sales), the flows of scrap material must be removed from the A matrix. A revised formulation can be illustrated with reference to the hypothetical example shown earlier. It embraces the following assumptions.

(a) Scrap is not actively produced but is a by-product.

(b) Scrap is only required by process P (billet) and s_1 units are required per unit output of P.

(c) Processes Q, R and S produce λ_2, λ_3 and λ_4 units of scrap per unit output of their principal product.

It follows from assumption (a) above that there are no multiplier effects on the production of scrap, and also that scrap is purchased from external sources to the extent of the short-fall

TABLE 4. THROUGH-GROUP COST STRUCTURES BY PRODUCT GROUP (STANDARD FORMULATION)

	Product P	Product Q	Product R	Product S
Scrap-external	0.2048	0.0506	0.0623	0.0332
Other material	0.0177	0.2513	0.1876	0.1291
Labour	0.4439	0.3565	0.4391	0.4791
Gross margin	0.3336	0.3416	0.3110	0.3586

TABLE 5. REVISED MODEL FORMAT

Sales from \ Sales to	Product group P	Q	R	S	External sales	Total sales
P (Billet)	—	200	200	—	100	500
Q (Rod)	—	—	200	300	300	800
R (Wire)	—	—	—	300	500	800
S (Fasteners)	—	—	—	—	1000	1000
Cost (external)						
Scrap	150	−10	−20	−20		
Other material	—	200	100	—		
Labour	200	200	200	250		
Margin	150	210	120	170		
Total	500	800	800	1000		

in within-group supply. The information in Table 3 can now be re-arranged so that scrap flows are located in the W matrix as negative inputs instead of remaining in the X matrix as positive outputs. Table 5 shows the revised layout.

Recalculating the multiplier effects of other processes within the system, unit demands for each of the process products leads to a row vector of scrap consumption, $\boldsymbol{\alpha}'$, where

$$\boldsymbol{\alpha}' = \mathbf{s}'(I - A)^{-1} \tag{6}$$

and $\mathbf{s}'$ is the row vector $(s_1\ 0\ 0\ 0)$.

Internal scrap supply by each process is given by the row vector $\boldsymbol{\beta}'$, where

$$\boldsymbol{\beta} = \lambda'(I - A)^{-1} \tag{7}$$

where λ' is the row vector $(0\ \lambda_2\ \lambda_3\ \lambda_4)$.

Therefore, external scrap supply for each process is simply the row vector $\boldsymbol{\gamma}'$, where

$$\boldsymbol{\gamma}' = \boldsymbol{\alpha}' - \boldsymbol{\beta}'$$
$$= (\mathbf{s} - \boldsymbol{\lambda})'(I - A)^{-1} \tag{8}$$

In the numerical example, the first row in the matrix $C(I - A)^{-1}$ gives the net external scrap supply requirement per unit output of each product group. The revised complete matrix is shown below as Table 6 and it is interesting to note that the 'through-group' gross margin of product P is now 30%. The initial formulation yielded a through-group gross margin on this product of 33.36%. In other words, it overstated the profitability of the product by about 11%.

The ability of the model to provide 'on-line' estimates of the net external scrap requirement for any given vector of sales to non-group customers, is a matter of considerable importance. This is particularly so where the amount of scrap material available for purchase from outside sources is restricted. Under these circumstances the ability to predict an imbalance between scrap required and scrap available will enable appropriate action to be taken well in advance of any potential shortage.

The cost matrix W (and hence C) can be extended to any number of rows which show additional effects. A row vector of working capital requirements could be introduced, for example, which would make it possible to compute the through-group working capital requirement per unit of sales from each company. In the same way, an additional row can be written into W containing the scrap production figures as positive entries. Let the new matrix be W^* with coefficient form C^*. The elements of W^* are shown in Table 7. The coefficient form can be used in the same way as before. The second row of $[C^*(I - A)^{-1}]$

TABLE 6. THROUGH-GROUP COST STRUCTURES BY PRODUCT GROUP (REVISED FORMULATION)

	Product P	Product Q	Product R	Product S
Scrap—external	0.3000	0.0625	0.0656	0.0184
Other material	0.0000	0.2500	0.1875	0.1313
Labour	0.4000	0.3500	0.4375	0.4863
Gross margin	0.3000	0.3375	0.3094	0.3640

TABLE 7. REVISED 'OUTSIDE-GROUP' PURCHASES MATRIX W*

Costs	Product group P	Q	R	S
Scrap	150	−10	−20	−20
Scrap (internal)	—	10	20	20
Other material	—	200	100	—
Labour	200	200	200	250
Margin	150	210	120	170

TABLE 8. REVALUED COST MATRIX $\overline{W}$

Costs	P	Q	R	S
Scrap	145	−9	−18	−18
Other materials	—	200	100	—
Labour	200	200	200	250
Margin	155	209	118	168

simply gives the total 'through-group' value of scrap generated per unit output of each product group, the second element of $[C^*(I - A)^{-1}\,\mathrm{B}\hat{\mathbf{y}}]$ yields the total quantity of scrap generated within the group for any given vector of sales $\hat{\mathbf{y}}$ to the nine main end markets.

A most interesting consequence of this approach to the scrap recycling problem arises from the valuation of the scrap in the context of the seller, the buyer, and the group as a whole. If process *P* had to purchase all of its scrap requirements from external sources, it would be paying (say) £40 per tonne. This would include a merchant margin of (say) £4 per tonne. On the other hand, if the scrap-producing companies within the group were forced to sell their scrap to non-group merchants, they would receive £36 per tonne. Thus, given that product group *P* has to consume scrap and product groups *Q*, *R* and *S* cannot avoid producing it, the group is better off by £4 per tonne as a result of being able to recycle its own scrap material. If *Q*, *R* and *S* are selling scrap at the full merchant price, then the entire benefit will initially be shown in their accounts. If the sales to *P* are made at £36 per tonne then the full benefit will appear in the costings of product *P*. Moreover, the benefit will vary over time since it is a function of the product mix, which itself will be changing, so the through-group cost structures will themselves be a function of the product mix being sold. The question naturally arises as to whether the benefit which stems from the saving of the merchant margin can be more appropriately distributed amongst the product groups.

The estimates of the total value of 'through-group' scrap produced, by product group, for any given vector of external sales can now be revalued to reflect the merchant margin retained within the group. This revaluation can be carried out by amending the matrix *W* (shown in Table 5) to allow for a 10% merchant margin. The revised elements are shown as matrix $\overline{W}$ in Table 8.

This amendment preserves the accounting balances of the original table and, of course, similarly revised matrices would be generated for each year over the time horizon for which the forecasts are being made. Dividing elements in the columns of $\overline{W}$ by original column totals yields the coefficient form $\overline{C}$. As before, $\overline{C}(I - A)^{-1}$ gives the final through-group cost structure by product group. This is given in Table 9.

It is noteworthy that a comparison of Tables 6 and 9, both of which are revised formulations of the original but with different scrap valuations, the through-group scrap costs are reduced in all four product groups and gross margins increased as a result of the revaluations.

4. CONCLUSIONS

This paper has attempted to develop some methodological issues underlying the application of input–output analysis at the micro level. It is a technique which has much relevance in the context of a vertically integrated

TABLE 9. THROUGH-GROUP COST STRUCTURE BY PRODUCT GROUP (REVALUED SCRAP)

	Product P	Product Q	Product R	Product S
Scrap	0.2900	0.0613	0.0653	0.0180
Other material	0.0000	0.2500	0.1875	0.1313
Labour	0.4000	0.3500	0.4375	0.4863
Gross margin	0.3100	0.3387	0.3097	0.3644

Vol. 6
1978

group of companies. Nevertheless the scrap recycling problem illustrates that appropriate modifications are both necessary and fruitful in serving a better analytical base for accounting and planning at group level.

The major benefit from this approach lies in the fact that there will be no need to reinvert the internal transactions matrix as a result of a sharply changing vector of external sales. This would have been unavoidable if the standard format illustrated in Table 3 has been retained. It is only necessary to revalue the C matrix and perform a limited amount of matrix multiplication, all of which can be accomplished very quickly. It therefore remains possible to maintain, at all times, an on-line immediate response system as an integral part of the planning process.

REFERENCE

1. CAMBRIDGE DAE (1963) *A Programme for Growth*, Vol. 3, Chapman & Hall, London.

ADDRESS FOR CORRESPONDENCE: *Jeffery I Round Esq., Department of Economics, University of Warwick, Coventry CV4 7AL, UK.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 515–522
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0515$02.00/0

The Teaching of Business Policy— A Fresh Approach

R J BETTS

Imperial College, London

(*Received July* 1978)

The artificiality of the circumstances in which much of management education takes place is exemplified in the teaching of business policy where the missing ingredient of responsibility is most crucial. The use of ready made cases compounds this problem by presenting the student with a partly identified problem and the information he must use in tackling it. A different approach in which students prepare their own studies of named companies and discuss their analysis and proposals with senior executives from those companies reduces to a worthwhile degree the artificiality of much business policy teaching. This paper discusses the problems of teaching in this field and describes an approach using the method described above.

THE STUDY of 'Business Policy' under various titles is a well-established component of most post-graduate business and management courses both in America and in Europe. It is often treated synonymously with corporate strategy or at least some aspects of corporate or long-range planning. The impression gained by this author from readings in the field and discussions with some of its teachers, is that despite its apparently firm place in management education, business policy is a subject which gives rise to an unusual degree of questioning and doubt as to its purpose, its intellectual substance, its methodology, and not least, as to the appropriate method of teaching it.

The rationale for the inclusion of the subject in management courses appears to be that there is undeniably a management function which corresponds to it in that all organisations need an overall guiding strategy by means of which the various departmental functions of the organisation can be integrated and given a common direction. There is even a case for saying, indeed it is often said, that merely the more precise definition of an area of responsibility called corporate strategy or business policy, quite apart from the substance of the topic, is one of the contributions to management in this area that management education can make.

The thinking behind this is that attention to the long-term direction of the company is likely always to appear to be less urgent than the pressing emergencies of day-to-day management. The definition of strategy is always postponable and unlike what Ansoff [1] defines as operating decisions does not automatically impose itself on management. Again as Ansoff points out, the temptation is ever present to regard short term and urgent problems as being unrelated to each other and amenable to correction without fundamental reappraisal of the company's general position and direction. In many, if not most cases, this view is justified but in many others the apparent problems of management detail are merely symptoms of an underlying and growing malaise which must ultimately and probably in less favourable circumstances demand of management a re-definition of the company's longer term direction and it is this which is the special contribution of business policy to management education.

Most teachers in the field insist that business policy is more than merely integrative in its function. That is to say, it does not merely pull together and co-ordinate the different departmental functions of the company. The definition of strategy concerns the direction of an integrated unit or at least of a unit which can

become integrated partly as a result of the process of strategy definition.

In this, great stress is placed on the interaction between the company and its external environment. The environment is conceived of principally in terms of actual and potential customers, competitors, suppliers, and government. Deriving mainly from the framework proposed by Ansoff, strategy definition then involves a search of the environment for threats and opportunities which developments will present to the company, appraisal of the tangible and intangible resources of the company often presented in terms of its strengths and weaknesses, the generation of alternative responses to developments in the environment and finally the selection of that response (i.e. that strategy) which promises the best match between the resources of the company and the threats and opportunities which are expected to arise (see also [3]).

THE TEACHING OF BUSINESS POLICY

So far we have described the managerial substance of business policy, that is the functions of management which correspond to what is taught under that heading. It may now be instructive to consider how the subject may best be taught in the light of suggested criteria for a good teaching method.

Even the use of the term 'teaching method' highlights one of the major problems—a problem which arises with respect to all the various subjects of management education in some degree but one which is perhaps exemplified by this particular instance. This is, of course, that the subject is being approached in a highly artificial context.

The principal artificiality is that the students (and the teachers) of the subject carry no responsibility for the implementation of anything they may propose and no responsibility either for the consequences of its implementation—or at least no responsibility which can be made in any way commensurate with the responsibility which a Chief Executive or his Board of Directors must carry, and of which they must always be conscious.

This is a crucial weakness. It is crucial because pre-eminently in this field, the uncertainties are overwhelming. The definition of corporate strategy can never be a matter of calculation. It is in this field that the recognized techniques of management science and operational research have made least headway. Such apparently quantitative models as have been proposed are based for the most part on subjective probabilities which are essentially means of expressing executive judgements and a working through of their possible consequences.

Because of the uncertainties which pervade strategy definition, it is all too easy for those who have no responsibility for implementation to propose strategies which have great but meretricious appeal in terms of their apparent boldness or perhaps to argue along a line simply as an exercise in persuasive advocacy.

It is, of course, an axiom of university education that students be encouraged to break away from stereo-typed thinking, that they should read and speculate without preconceptions and be able to generate and express new perspectives without fear of incurring the disapproval of their teachers. Something of this spirit should be encouraged in the teaching of business policy and is allowed for in the framework of analysis we have described. Similarly, the climate of management and the procedures of analysis should allow a free-ranging debate in strategy definition within the company. However, something of the sobriety which may be expected to inform the process of strategy choice within companies should be introduced also into the artificial circumstances in which management education inevitably takes place. The potential for introducing discipline and responsibility into the learning process must be one of the first criteria of a good teaching method in this field.

A second criterion relates to the way in which the method and the materials it uses present the task with which the student is faced.

As we suggested earlier, the manager is not, in most instances, presented with the problem of defining corporate strategy. This is a problem he identifies and defines for himself. His circumstances present him with responsibility for a company which he perceives perhaps in terms of people, tangible assets, production processes, raw materials, markets, financial data, responsibilities, and so on. He is not pre-

sented with a readily identified and defined problem relating to the company's longer term strategy. It requires an initiative on his part to distinguish from his perceptions of the business what his strategic problems are.

Similarly, business policy should not be taught in such a way that the students are presented with a defined task in the form of an identified strategic problem which they are then expected to solve. It is necessary, of course, that at some point they have described to them in general terms what the concept of corporate strategy is and be given illustrations of strategy-level problems. They may also be taught a framework of analysis, perhaps of the Ansoff type so that they may, if they choose, at least begin by applying that framework. The important point is that when they come to examine a particular company, the information they are given about that company should not be so structured that 'the problem', that is any problem which seems most appropriate to the teacher, is defined for them. The senior executive is able to call for any amount of information he is prepared to pay for and by his choice of the information he requests, he expresses something about his perceptions of the task before him. However, the responsibility and the choice are his and he can accept no-one's instruction as to the nature of his task. It is appropriate that as far as possible the design of the teaching method places the student initially in a similarly unguided position.

A further criterion relates to the behavioural context in which decisions about strategy are reached. In terms of formal authority the chief executive may be empowered to determine strategy. In reality, if only because he needs to carry his colleagues along with him, the formulation of strategy will involve a process of debate much, if not all of it, oral and a movement towards a consensus.

In this, the familiar skills of consultation, advocacy and debate will be crucial and the strategy which emerges will reflect the interactions and the compromises which have taken place. Thus, at this senior level, the definition of strategy is not a process only of the intellect carried out behind executive desks. Again, since the teaching and learning process should mirror as far as possible the reality, students should not tackle their problems in isolation or be allowed to suppose that they solve their problems by convincing themselves alone of the correctness of their solution.

It was suggested earlier that uncertainty is pervasive in strategy formulation and that this carried the danger of indiscipline in the student's approach to problems. This has a further implication for teaching, namely that the necessity to make decisions in the absence of complete information must be brought home to the student. Imaginative students are usually skilled in presenting alternatives with their advantages and disadvantages. Again in many respects this is a desirable product of their education, i.e. the facility to present a balanced view based on a careful and objective assessment of evidence. However, in management education this is not sufficient and the student should not be allowed to conclude only that he would need more information in order to reach a conclusion, and thus to make a firm recommendation. This may be sufficient for someone in one of the management services who can present evidence and analysis and then leave to his 'client' in management the responsibility for choice. This implies that it should be a necessary part of the student's task to present firm statements of the policies he proposes to adopt despite the partial nature of the information on which he has been obliged to base his decision. This is, after all, the very situation in which the executive finds himself and the responsibility of the teacher is to reflect the reality as far as possible even if, as in this case, this goes against the grain of much of the student's previous training.

IMPLICATIONS FOR TEACHING METHODS

It is probably worth making the comment even at this stage, that the term 'teaching method' is less appropriate in this context than the term 'learning method' as a description of the process whereby educational objectives are met. Even those in management education who are not in general proponents of the case method seem to feel that the method of 'learning by doing' should occupy the greater part of the student's effort in the business policy field. The following remarks, as well as the foregoing, are based on that view.

The discussion above suggests that the educational method should attempt to achieve the following:

(a) the student should be made to feel responsibility for his proposals.

(b) he should not be presented with identified and structured problems.

(c) the method whereby proposals are reached should involve collective effort and the process of debate.

(d) the student should be obliged to make firm proposals for action despite inadequate information.

Perhaps the most widely used text for business policy courses (apart from 'framework of analysis' texts such as that by Ansoff mentioned earlier) is the set of cases produced by the Harvard Business School [2]. Whilst the authors of these cases would probably not disagree with much of what has been argued here, in the view of the present author the ability of any set of ready-made cases to meet the proposed criteria is inherently limited.

A skilled teacher using prepared cases may doubtless be able to conduct a disciplined discussion in which the rash and the frivolous are constrained by the norms of the class or the wit of the teacher. However, both students and teacher know that their involvement with that company is confined to the classroom, will soon be over, and does not carry the threat of public examination or of serious damage to their career at least within very wide ranges of performance. Within an educational context there is, of course, no practicable means whereby the nature of the executive's responsibility can be reproduced. Nevertheless it may be possible to impose additional and rather different penalties on the student for faulty analysis or advocacy than simply the derision of his classmates or the reproof of his teacher.

The second criterion of the degree of structure is usually recognized in the prepared cases by the absence of set questions for the student to answer and by the presentation of a mass of information about the company without apparent indications of relative importance or overt classification. It is true that in the book of Harvard cases, the cases are placed in sections and with accompanying text which have headings which indicate the broad character of the situations they represent. It is likely that this will predispose the student to look for aspects of each case which correspond to the topics covered in the text but nonetheless the cases are not clearly structured in such a way that the student is closely constrained in what he may diagnose to be the problems which the company faces and to which he, therefore, feels he must address himself.

It remains inevitably true, however, that the case writer imposes structure on the problem by virtue of what he chooses to include in his account of the company and what he chooses to exclude. Certainly all cases include some amount of redundant information and, whether justifiable or not in the eyes of the case writer, almost all of them leave the student wishing for additional information on some aspects of the company or its environment. Certainly the student finds himself in a very different situation from the executive who begins his task of problem definition and strategy formulation with no such helpful collection of putatively relevant information.

When we turn to the third criterion, the necessity for group interaction and collective decision, there is no reason why the use of written cases should not be organized in such a way that this is achieved. However, the value of this should not be divorced from the first criterion concerning the pressure of responsibility. Much of the urgency and the heat disappears from debate and collective decision making if nothing of great personal consequence is thought to rest on the outcome. In such a case, winning the argument may come to be more important to the participants than getting the 'right' answer. The ability to win arguments may indeed be a crucial managerial skill but the opportunity to exercise it for all but a short time requires that it be deployed at least some of the time in support of sound proposals. It is only when the participants have an awareness of significant penalties following from proposals which are unsound that the process of reaching those proposals bears any resemblance to the corresponding process in management.

In a similar fashion there is no reason why the use of written cases should not require the

student to produce firm proposals for action supported by evidence and argument. Here again, however, this may be relatively lightly undertaken, despite a lack of complete information, if subsequent events exact no penalty for bad proposals.

What is being argued here, then, is that on a number of grounds a business policy course should not come to be regarded in a purely academic light. A sense of involvement with the company and a feeling of responsibility or even trepidation in strategy formulation is, if not an absolute necessity for success, then certainly an immensely important contributory factor and one worth going to considerable trouble to achieve even if necessary at the expense of some other and more conventional teaching objectives.

BUSINESS POLICY AT IMPERIAL COLLEGE

The Department of Management Science at Imperial College has operated a Master's programme under various guises for over two decades. For the last twelve of those years an elective course in Business Policy has been offered usually attracting 30–40 students.

The design of this course has been deliberately formulated with a view to minimizing the deficiencies which were perceived in the use of pre-existing case studies. As far as is known, the design which has evolved is different in essential respects from that used elsewhere, and whilst the present author is not unaware of its own peculiar shortcomings, our experience with the design leads us to think that it has many advantages over more conventional approaches.

After a brief introductory set of lectures which describes the purpose and objectives of the course and suggests an appropriate framework for the approach to corporate strategy, the students are divided into groups of about five members. Each of these groups is then assigned to a company. As far as is practicable the students form their own groups and choose their own company from the set provided. The companies concerned have previously been approached usually in the person of the Chief Executive or Chairman and their cooperation obtained. The choice of companies to approach is usually made by the two teachers who are most concerned with the organization of the course and the choice is made usually because the company's strategic position offers some particular interest. It may for example have gone through a period of especially rapid growth or be faced with some kind of major change in its markets or in related technology. On the other hand it may be of interest e.g. because although in a relatively stable position it is smaller than its rivals or conversely because it is sufficiently close to being a monopoly for this to be of interest. The students are not told initially the factors which made the company of interest although they may be told subsequently in the small minority of instances where this does not occur to them spontaneously.

Over periods ranging from two to six months depending on timing within the course, each group then prepares a report on its company of up to 20,000 words. This presents an analysis of the company, its performance (as far as possible in comparison with its closest competitors), and of developments in its environment. In the final section the students are required to set out the strategy which the company should follow in meeting the problems and opportunities which they have themselves identified as being important. They are not, as a matter of normal practice, presented with information by their teachers but only with suggestions of possible sources of information. The students are themselves responsible for finding the information they require. Some of these sources are those which are publicly available such as newspapers and periodicals, but some require the students to make contact with e.g. stockbrokers, trade associations, government departments, or rival companies in the search for information and opinion.

Only when such sources as these have been exhausted will the students approach the companies themselves. The purpose of this approach is for the students to ask for any information which they have been unable to obtain elsewhere and which they believe is important to their analysis and to resolve any issues of interpretation about which they feel in doubt.

An individual copy of their report is then given to each member of the class. This is followed some days later by a preparatory

seminar in which the students as a group present their report to the class and lead a discussion on the issues it raises. The purpose of this seminar is to provide an opportunity for the class to challenge the analysis and the proposals which the group have set out and for the group to defend them. As a result of this process the class develops a sharper perception of the company and its situation and in some cases, though by no means in all, moves towards a common view.

In the meantime, copies of the students' report have been sent to the Chief Executive or Chairman of the company concerned and about one week after the preparatory seminar a second seminar is held which is addressed by (in almost every case) the Chief Executive or Chairman of the company usually accompanied by one of his colleagues, often his Finance Director or senior manager responsible for planning.

At this seminar it is the company visitor who plays the leading role. Its purpose is to allow the company in the person of its most senior manager to react to the view of it and its proper direction which the students have presented, and to present his own view as to whether the questions they have raised are the important ones, whether their analysis is correct and their proposals valid. The seminar is regarded as being confidential (as is the report) and the visitor is encouraged to speak as frankly as he is able both about his company and about the students' report.

One obvious danger in such a procedure is that the visitor will wish to spare the feelings of the students, even if his reaction is adverse, and his remarks will be correspondingly anodyne. This does sometimes happen but the students are in no position to predict whether it will in their own case and they see sufficiently numerous examples of outspoken comments that they are always unwilling to rely on it in producing an insufficiently researched or irresponsible report. Normally, the prospect of facing someone of that standing in a company they have come to know and feel involved in, who will have had an opportunity to study their report in depth and to refer the whole or part of it to specialist colleagues, is quite sufficient to induce at least something of the seriousness which must always attend high level policy making in the real world.

The seminar itself is followed by an informal small dinner party attended by the visitor or visitors from the company and by the students responsible for the report.

Although social in atmosphere the purpose of this is to enable the students to have further discussions with the visitor and thereby to gain further insights into his thinking and into the problems of management at the most senior level. In this they are helped immeasurably by their knowledge of the specific circumstances of the company.

In any year six companies will normally be studied and during the time for which the course has been in operation, sixty-five companies in all have been involved (some studied more than once over the 12 year period in question).

THE MERITS OF THIS APPROACH

We believe that, whilst not without its drawbacks, this approach to the study of business policy comes much closer than does the traditional use of written cases to meeting the criteria suggested earlier. The students are presented merely with an identified company rather than with an identified problem to solve. The exercise is more real and tangible because they know that the company is real, the issues live, and that they will have direct contact with at least one of the principal characters in the drama that they are studying. They also know that their efforts will be judged by one who is in a far better position to judge than they are themselves. They produce the report as a group and they must justify it in debate with their classmates and teachers. They are obliged to make proposals for action on the basis of information they know to be incomplete and with a degree of seriousness which the context of the exercise requires.

Moreover, there are other subsidiary but important benefits following from this approach. Firstly, the students have to produce a report which quite apart from the quality of its content must meet the criteria of good report writing. That is to say it must be logically structured, literate, lucid, appropriately referenced and generally persuasive. Since most of our students are from an undergraduate and industrial background in engineering, science, or

mathematics, this is an experience which to most of them is new and in which they need experience and guidance.

Secondly, it introduces the student to the sources of data and the techniques of research into companies and markets and to some of the practical difficulties and shortcomings which they involve.

Thirdly, and in some ways most importantly, it gives the students an opportunity to meet and debate with executives at the most senior level in an informal environment. This for most of them is likely to be the first and last occasion they have for this kind of contact. Subsequently they will view such people as superiors, employers, or perhaps competitors. Thus they have what is probably a unique opportunity to gain an objective impression of the outlook and the attitudes of someone in that position.

There are, of course, disadvantages in this approach. One of the major ones is that the exercise is very time consuming and this therefore limits the number of companies that any one class of students studies. The use of book cases for the same input of time and effort would bring the student into contact with a much greater number of separate companies. The contact would of course be much more superficial and much less real. It is an important matter of judgement to decide whether more examples in less depth are superior educationally to fewer examples in greater depth. As the approach we take would suggest, we believe for the reasons I have set out, firmly in the latter alternative. The author's view is that taking the class through a larger number of cases is superior in giving more opportunity for the development of a general analytical framework. The students certainly become more skilled thereby in the analysis of book cases but it remains to be shown that they also become more skilled in the analysis of companies. The proponents of this method would doubtless argue that this is one of the advantages, namely that it clears away the detailed specifics of the company to reveal the essential strategic problem. Against this it may be argued that a written case does not present the many facets of the specific circumstances of the company in a sufficiently real way for this to be a valid experience of the required analysis.

A second disadvantage is that the method of report writing makes it difficult to study anything other than large publicly quoted companies. This is simply because smaller or unquoted companies will be impossible to investigate adequately without what may be an undesirable degree of reliance on data and comment supplied by the company itself. Nevertheless we have from time to time successfully included small companies in our programme where the students have been given ample access to necessary data and comment from the company which has been frank and objective.

The other dimension, the spread of industries represented has been very successfully wide, ranging from merchant banking through retailing to engineering, chemicals, and construction. The readiness of most companies to take part and the openness of the visitors in the seminar discussions have been most encouraging. On the whole the smaller companies have been most free in what they are prepared to reveal and those which are most closely involved with government, perhaps because they are also most in the public eye, have been least free. Many executives have confessed their surprise, and perhaps even alarm at how much of their business can be revealed from intelligent use even of only published information. One senior executive of a large company agreed to take part but on the condition that no approach be made to his company precisely because he wanted to see what could be discovered and what impression would be formed from public information and comment.

CONCLUSION

Our experience with this course encourages us to believe that, as an approach to the teaching of business policy, it has much to commend it compared to the more traditional approach through written cases. The reactions we have had from the executives who have taken part have been very sympathetic to the form of the exercise and in most though not every case, favourable to the report which the students have produced on their company. In addition, past students a few years out in their careers have rated this course as the most worthwhile of all in spite of, or perhaps in part because of, the considerable work load which its design entails.

One final point may be worth making. Almost without exception, the executives who have participated in these seminars have shown an ability to deal with their situation and policies on a conceptual as well as a concrete level, and have been able to 'verbalize' their thinking with a clarity which would shame many professional teachers. Whatever problems the design of this course has encountered the quality of the communicative skills of its participants from industry has not been one of them.

REFERENCES

1. ANSOFF HI (1965) *Corporate Strategy*. McGraw-Hill, New York.
2. LEARNED EP, CHRISTENSEN CR, ANDREWS KR & GUTH WD (1965) *Business Policy: Text and Cases*. Richard D Irwin, Homewood, IL.
3. MCCARTHY DJ, MINICHIELLO RJ & CURRAN JR (1975) *Business Policy and Strategy*. Richard D Irwin, Homewood, IL.

ADDRESS FOR CORRESPONDENCE: *RJ Betts, Esq., Department of Management Science, Imperial College, Exhibition Road, London SW7 2BX, UK.*

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 523–530
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0523$02.00/0

A Nuclear Medicine Patient/Instrument Scheduling Model

JAMES D JOHANNES

RICHARD M WYSKIDA

The University of Alabama in Huntsville, USA

(*Received June* 1978)

The problem of scheduling patients and clinical instruments for each physician-requested Nuclear Medicine study, in a clinical environment, is resolved through the development of a computerized heuristic model. The study schedule was determined by minimizing such factors as elapsed instrument time, instrument idle time, and maximizing instrument utilization. The heuristic scheduling procedure was developed and evaluated for scheduling thirteen different Nuclear Medicine study types on a daily basis. The analysis showed that this heuristic can be utilized to provide a good basic schedule for use in Clinical Nuclear Medicine.

INTRODUCTION

RAPID and extensive technological advances in modern medicine have significantly reduced the death rate in our nation and have forced hospital administrators to seek more effective use of both personnel and equipment. The need to interact with many widely specialized medical professionals, together with efforts to reduce costs through effective computer utilization to augment decision-making, has had a major impact on these health organizations. This paper examines only one of the many problem situations in medical facilities, that is, the problem of scheduling patients and clinical instruments for each physician-requested Nuclear Medicine study.

The first section is devoted to the establishment of the clinical scenario. Section II examines the development of a scheduling heuristic for a Clinical Nuclear Medicine facility. Finally, the results of an example problem, together with validation experience, are discussed.

THE CLINICAL SCENARIO

Research and practicing physicians have turned to the many disciplines outside of medicine for methods and techniques in their search for more effective means of diagnosing and treating their patients. In order to provide a better service to their patients, many physicians have specialized in such fields as cardiology, orthopaedics, obstetrics, urology, and radiology. Many of these physicians are associated with hospitals or clinics because of the large investment required for specialized equipment, the proximity of research being performed, or for the simple reason that hospitals and clinics provide the proper environment for specialized health care.

The radiology community found itself confronted with a massive and growing information processing load in the early 1960's. The principal contributor to this growth was a relatively new technology within radiology, that is, Clinical Nuclear Medicine.

Clinical Nuclear Medicine may be divided into two major areas: *in vivo* and *in-vitro* studies. In the *in-vivo* study, a radioisotope is injected into the patient and the clinical procedure is performed using a single scanning instrument (machine), resulting in a nuclear scintigram. The *in-vitro* study requires the drawing of a small amount of body fluid from the patient and a chemical assay test performed.

The three basic types of studies performed are the static probe, dynamic function, and imaging study [5].

The radiation count is sampled at varying times, depending upon the type of study being performed. For example, in the case of a brain bolus and the accompanying brain scan, the bolus requires immediate scanning during isotope injection and then again after approximately three hours have elapsed to determine the steady state condition. Since the counting of these nuclear events may be easily digitized, computerization of the data gathering and display aspects of Clinical Nuclear Medical is possible. Scheduling, performing, evaluating, and reporting these studies, while maximizing the physician's efficiency, presents a continuous challenge to the Nuclear Medicine division personnel.

With these thoughts in mind, the administration of the Nuclear Medicine Division of the Vanderbilt University Hospital determined that the role of the computer in the clinic would be expanded. Primary emphasis would be upon the "automation" of instrument scheduling.

Instrument scheduling expedites the patient's progress through the facility and also provides the proper sequencing of related studies to insure more accurate study results. Such scheduling permits the physician to achieve a high level of time utilization.

The Nuclear Medicine Division of Vanderbilt University Hospital possesses the capability to conduct 13 different Nuclear Medicine study types utilizing five different instrument combinations. The schedulable instrument study processing time, radioisotope injection time, instrument setup time, and study priority for each study type are summarized in Table 1. Each study type requires only one instrument, or in some cases a combination of an instrument and the Intertechnique high speed data collection device, with the data processed at a later date. The priorities in Table 1 are derived from the following study constraints:

(1) The lung studies are considered emergency studies. A request for this study normally occurs when the patient suffers from impaired respiration.

(2) There is only one Intertechnique data collection system.

(3) The Brain Scan Study must be performed between 120 minutes and 240 minutes after completion of the Brain Bolus.

The thirteen study types are thus partitioned

TABLE 1. STUDY TYPES AND RELATED VARIABLES

Study type (*i*)	Schedule symbol	Schedulable instrument* (*m*)	Study processing time (min) (ST)	Radioisotope injection time (min) (IT)	Instrument setup time (min) (ET)	Study priority
1. Brain bolus	BB	4,5	15	5	5	2
2. Brain scan	BS	1,2,3	60	0	5	2
3. Cardiac scan	CI	4,5	30	5	5	2
4. Cisternogram (4 elapsed hours)	C1	1,2,3	30	5	5	3
5. Cisternogram (24 elapsed hours)	C2	1,2,3	30	0	5	3
6. Cisternogram (48 elapsed hours)	C3	1,2,3	30	0	5	3
7. Liver–spleen scan	LV	1,2,3	60	5	10	3
8. Lung scan	LS	1,2,3	30	5	10	1
9. Lung scan—XENON	LX	5	30	5	15	1
10. Renogram scan	RI	4	30	5	5	2
11. Renogram—bilateral	RB	4	30	5	10	2
12. Renogram triple	RT	4	45	10	10	2
13. Thyroid scan	TY	4,5	60	5	15	2

* The following notation identifies the various instruments:
1. High Resolution Anger Camera No. 1.
2. Low Resolution Anger Camera.
3. High Resolution Anger Camera No. 2.
4. High Resolution Anger Camera No. 1 and Intertechnique Data Collection System.
5. Low Resolution Anger Camera and Intertechnique Data Collection System.

into three priority classifications:

(1) Emergency.

(2) Intertechnique data collection system requirement or a study that is time dependent on a study requiring the Intertechnique system.

(3) No special requirements.

These classifications serve as the general sequence for assigning the studies to the instruments in order to satisfy the constraints.

SCHEDULING HEURISTIC

The problem of scheduling or sequencing n patients on the five instrument combinations is a time-consuming task. The basic problem is to minimize the total time to perform the physician-requested studies in the clinic, thus minimizing equipment idle time during the physician's prime time in the hospital. In this environment, a desirable solution would incorporate the following elements:

(1) Study type.

(2) Instrument requirements.

(3) Instrument setup time.

(4) Study or scan time.

A common approach to scheduling problems is the employment of a heuristic procedure. A heuristic is a procedure that has been found to provide satisfactory solutions to a problem, but it may not be optimal and does not possess mathematical proof at the present time. An example of a heuristic would be to process first the job with the shortest job time and the closest deadline, with the goal in mind of minimizing the number of jobs that fail to meet the deadline.

A special case of the scheduling problem, in which each job consists of a single operation on a single machine, can be applied to clinic patient scheduling problems since each study can be considered a single operation. Each set of studies (jobs) can be partitioned, depending on the instrument or instrument combination (machine) required to perform the study. Consequently, the developed heuristic will be single instrument oriented.

The total time for each study, on the schedulable instruments is $(P_{i,m})$ and N_i is the number of patients per study type. The total block time is therefore,

$$B_{im} = N_i P_{i,m} = N_i(ST_{i,m}) + \text{Max}\{IT_i, ET_{i,m}\}.$$

Since several study types require the Intertechnique Data Collection System, and since it may be connected only to Instrument 1 or Instrument 2, idle time may have to be inserted on these systems. Idle time may also be a function of the particular study constraints, as in the minimum waiting time between a brain bolus and a brain scan. The idle time may be minimized by selecting the instrument with the minimum amount of time elapsed, as long as its satisfies the study constraint. Therefore, the total elapsed time for an instrument may be expressed as

$$TM_m = \sum_{i=1}^{13} B_{im} + ID_m.$$

Thus, the objective function for the patient study/instrument scheduling problem is defined as

$$\min F = \sum_{m=1}^{5} TM_m.$$

To provide an optimal solution to the objective function, all possible permutations of the schedulable sequences would have to be evaluated and the optimum schedule selected. The shortest-processing-time (SPT) single machine theorem can be utilized within each priority classification since the processing time can be computed for each schedulable instrument. The optimality of SPT sequencing is established by the following [2]:

The mean-flow time is minimized by sequencing instruments in order of nondecreasing processing time (ST) as

$$ST_1 \le ST_3 \le \cdots \le ST_6 \le ST_8 \le \cdots \le ST_{12} \le ST_2 \le ST_7 \le ST_{13}.$$

If the SPT rule is applied to priority classifications 2 and 3 separately, a minimum sequence is assured for that instrument. If all selected instruments possess the least assigned time, a minimum total time for the instrument set is assured. Thus, the constraints are satisfied and

the objective function possesses a minimal value.

The heuristic procedure may be summarized into four separate steps as shown below:

Step

1. Construct the maximum block time B_{im} matrix.

2. Assign the emergency studies.

A. Assign study type 8 to instrument 3.

B. Assign study type 9 to instrument 2 and the Intertechnique Data Collection System.

3. Assign Intertechnique-oriented studies in the following order:

A. Study type 1 to instrument 1 and Intertechnique or instrument 2 and Intertechnique.

B. Study type 2 to instrument 1, 2, or 3 allowing for proper time interval between a bolus and scan (minimum of 120 minutes, maximum of 240).

C. Assign balance of studies (types 3, 10, 11, 12, and 13) requiring the Intertechnique instrument according to the following rule:

The study which has the least time requirement is assigned to the instrument combination which has the least total elapsed time.

4. Assign the remainder of the studies to the instrument having the least total elapsed time. Use the shortest processing time rule (SPT) for study selection. Unblock studies, if necessary, to fill idle time.

A discrete stochastic simulation modeling method was chosen to test the scheduling heuristic. The time period used in the simulation is one day since approximately ninety-five percent of the schedulable studies are known at the beginning of the work day. Varying periods of time, from one day to a full five-day work week, were simulated, with a restart for each day.

The method of simulation utilizes random numbers to represent the stochastic variables, which in this case are the number of individual patients per study type per day. Sixty-three working days of data, by study type, were collected from the clinic. Cumulative probability distribution functions (CDF's) were calculated, to four decimal places, permitting random numbers to be associated with the number of patients per study per day. The random number generator routine, adopted from Emshoff [3], is a multiplicative congruential method, which is the most commonly used method in simulation models. It is assumed that the empirical distributions are adequately approximated by the simulation procedure. The initial seeds were computed by using the commonly utilized reduced Fibonacci sequence

$$S_i = S_{i-2} + S_{i-1} \pmod{m} \qquad (7)$$

where $S_0 = 0$ and $S_1 = 1$ and m is 10^6, which results in a five-digit seed [1, 3]. The instrument setup, injection, and processing times are average values provided by the head technician in the clinic and were shown in Table 1. The unit of time used in the simulation was five minutes.

Three different heuristic procedures were programmed and analyzed in the process of developing the final heuristic procedure. The first procedure involved four classes. The added class considered study types 1, 2, 10, 11, and 12 as special cases. The second version combined the above special cases, with the exception of study type 2, with those studies requiring the Intertechnique system. The results showed that the special cases were not required. However, the time constraint for study type 2 was violated. The third and final heuristic considered study type 2 as part of class two, assigned immediately after study type 1, rather than class three. The assignment of the Brain Bolus and Brain Scan studies, and the use of idle time, underwent the most changes in the heuristic since the decision logic proved to be complex due to the time constraint and recognition of available idle time.

Table 1 defines the schedule table symbols used in the computer simulation and the scheduling program. The schedules are presented in a symbolic format by instrument. A four-character symbol denotes the study type and the number of patients in the block, "**" denotes idle time, and "ET**" denotes the instrument setup time.

After the heuristic procedure and the simulator model had reached a point of stability, that is, hospital personnel were unable to devise a superior schedule for several days of

Vol. 6
1978

input, the program was modified to accept a given patient workload for each study type. This modification allowed the model to be further tested by removing the dependence on the random selection process, and instead, accepting an actual clinic daily work load.

AN EXAMPLE

To illustrate the functioning of the heuristic procedure, a typical demand day was selected from the Nuclear Medicine Division records. Table 2 presents a summary of this information. Table 3 is the computerized heuristic procedure derived instrument schedule.

VALIDATION

The real test for the scheduling heuristic occurs when several days of actual clinic sche-

TABLE 2. NUMBER OF PATIENTS AND DISTRIBUTED PROCESSING TIME MATRIX [4]

Number of patients per study type					
Type 1	2	Type 2	2	Type 3	1
Type 4	0	Type 5	0	Type 6	0
Type 7	1	Type 8	1	Type 9	1
Type 10	0	Type 11	2	Type 12	0
Type 13	2				

Distributed time matrix				
1	35	35	0	35
2	125	125	125	0
3	35	35	0	35
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	70	70	70	0
8	40	40	40	0
9	45	45	0	45
10	0	0	0	0
11	0	70	0	0
12	0	0	0	0
13	135	135	0	135

TABLE 3. INSTRUMENT SCHEDULE

	Instrument			
Time	1	2	3	Intertechnique
800	**	ET**	ET**	ET**
805	**	ET**	ET**	ET**
810	**	ET**	LS 1	ET**
815	**	LX 1	LS 1	LX 1
820	**	LX 1	LS 1	LX 1
825	**	LX 1	LS 1	LX 1
830	**	LX 1	LS 1	LX 1
835	**	LX 1	LS 1	LX 1
840	**	LX 1	ET**	LX 1
845	ET**	**	ET**	ET**
850	BB 2	**	LV 1	BB 2
855	BB 2	**	LV 1	BB 2
900	BB 2	**	LV 1	BB 2
905	BB 2	**	LV 1	BB 2
910	BB 2	**	LV 1	BB 2
915	BB 2	**	LV 1	BB 2
920	**	ET**	LV 1	ET**
925	**	CI 1	LV 1	CI 1
930	**	CI 1	LV 1	CI 1
935	**	CI 1	LV 1	CI 1
940	**	CI 1	LV 1	CI 1
945	**	CI 1	LV 1	CI 1
950	**	CI 1	**	CI 1
955	**	ET**	**	ET**
1000	**	ET**	**	ET**
1005	**	ET**	**	ET**
1010	**	TY 2	**	TY 2
1015	**	TY 2	ET**	TY 2
1020	**	TY 2	BS 2	TY 2
1025	**	TY 2	BS 2	TY 2
1030	**	TY 2	BS 2	TY 2
1035	**	TY 2	BS 2	TY 2
1040	**	TY 2	BS 2	TY 2

TABLE 3—*continued*

Time	Instrument 1	2	3	Intertechnique
1045	**	TY 2	BS 2	TY 2
1050	**	TY 2	BS 2	TY 2
1055	**	TY 2	BS 2	TY 2
1100	**	TY 2	BS 2	TY 2
1105	**	TY 2	BS 2	TY 2
1110	**	TY 2	BS 2	TY 2
1115	**	TY 2	BS 2	TY 2
1120	**	TY 2	BS 2	TY 2
1125	**	TY 2	BS 2	TY 2
1130	**	TY 2	BS 2	TY 2
1135	**	TY 2	BS 2	TY 2
1140	**	TY 2	BS 2	TY 2
1145	**	TY 2	BS 2	TY 2
1150	**	TY 2	BS 2	TY 2
1155	**	TY 2	BS 2	TY 2
1200	**	TY 2	BS 2	TY 2
1205	**	TY 2	BS 2	TY 2
1210	**	ET**	BS 2	ET**
1215	**	ET**	BS 2	ET**
1220	**	RB 2	**	RB 2
1225	**	RB 2	**	RB 2
1230	**	RB 2	**	RB 2
1235	**	RB 2	**	RB 2
1240	**	RB 2	**	RB 2
1245	**	RB 2	**	RB 2
1250	**	RB 2	**	RB 2
1255	**	RB 2	**	RB 2
1300	**	RB 2	**	RB 2
1305	**	RB 2	**	RB 2
1310	**	RB 2	**	RB 2
1315	**	RB 2	**	RB 2
1320	**	**	**	**
1325	**	**	**	**
1330	**	**	**	**
1335	**	**	**	**
1340	**	**	**	**
1345	**	**	**	**
1350	**	**	**	**
1355	**	**	**	**
1400	**	**	**	**
1405	**	**	**	**
1410	**	**	**	**
1415	**	**	**	**
1420	**	**	**	**
1425	**	**	**	**
1430	**	**	**	**
1435	**	**	**	**

dules are compared with the heuristically derived schedules. Actual instrument times, idle times, and study schedules (including the number of patients by study type) were collected for each day for a five-day period. Table 4 contains the categories of actual clinic instrument times for each day. An asterisk indicates that the instrument was used beyond the normal work day. Table 5 is a summary of the heuristic developed instrument schedule time for the same five-day period.

Two measures were applied in comparing the generated schedules against the clinic's schedules. The first was a comparison of the total elapsed time by instrument for each day. Table 6 presents this comparison. Positive

TABLE 4. CLINIC WORK WEEK INSTRUMENT TIME (MINUTES)

Day		Instrument 1	2	3	Intertechnique
1	Inst.	445	375	430	285
	Idle	0	195	0	285
	Total	445	*570	430	*570
2	Inst.	545	435	560	610
	Idle	85	175	30	0
	Total	*630	*610	*590	*610
3	Inst.	370	260	315	275
	Idle	0	130	75	115
	Total	370	390	390	390
4	Inst.	280	135	185	230
	Idle	50	240	145	145
	Total	330	375	330	375
5	Inst.	460	240	380	325
	Idle	0	85	80	0
	Total	460	325	460	325

* Instrument time scheduled beyond normal work day.

values indicate the generated schedule improvement, whereas the negative values indicate that the actual clinic schedule was superior for that instrument. The second is the number of instrument days that each schedule version exceeded the normal working day.

The comparison of total elapsed instrument times showed a total of 1,145 minutes, or better than two instrument days, could have been saved by following the heuristic generated schedule. The first day showed the most improvement due to a better schedule of those studies requiring the Intertechnique data collection system. Part of the improvement can be attributed to the early scheduling of brain bolus studies, especially in days three and four.

The actual clinic schedule had six occurrences of instrument use past the end of the

TABLE 5. HEURISTIC DEVELOPED INSTRUMENT SCHEDULE TIMES (MINUTES)

Day		Instrument 1	2	3	Intertechnique
1	Inst.	385	460	405	280
	Idle	0	0	0	0
	Total	385	460	455	280
2	Inst.	475	565	435	605
	Idle	45	40	80	0
	Total	520	605	515	605
3	Inst.	370	220	355	275
	Idle	10	55	40	0
	Total	380	275	395	275
4	Inst.	280	135	185	230
	Idle	40	95	150	0
	Total	320	230	335	230
5	Inst.	460	170	445	325
	Idle	0	155	10	0
	Total	460	325	455	325

TABLE 6. ELAPSED TIME IMPROVEMENT (GENERATED VERSUS ACTUAL) (MIN)

	Instrument				
Day	1	2	3	Intertechnique	Total
1	60	110	−25	290	435
2	110	5	75	5	195
3	−10	115	−5	115	215
4	10	145	−5	145	295
5	0	0	5	0	5
Total	170	375	45	555	Σ 1145
Mean	34	75	9	111	229

normal work day. Instrument one was required 90 minutes past the normal work day on day two. The generated schedule had two occurrences, each 60 minutes, of instrument use past the normal work day. The actual schedules showed corresponding 90 and 70 minute requirements on day two. The generated schedule for this day showed an overall improvement of 195 minutes compared with the actual schedule. The mean values of improvement per instrument are also given in Table 6.

The generated schedules serve as a guideline and not as a hard and fast rule. The user must remain responsive to emergencies or other unplanned events. Given the basic schedule, the user is cognizant of idle times within the schedule and may use these times to accommodate emergencies. Alternatively, if idle time does not exist, the heuristic may be utilized as a dynamic scheduler to schedule the emergencies into the daily schedule.

The model may be extended to consider additional studies and instruments by adding the appropriate data and varying the array dimensions and DO-loop limits of the computer program. The current configuration of the model requires a total of 9,600 words of storage on the UNIVAC 1100/10. A daily schedule is produced in less than two minutes of elapsed time, including printing, using approximately two seconds of Central Processing Unit time.

CONCLUSIONS

This paper presents the development of a nuclear medicine patient/instrument scheduling model which has the capability of assessing thirteen different nuclear medicine study types that are assignable to five different instrument configurations. The heuristic scheduling program requires a minimum of input. The user need only supply the number of patients to be scheduled for each study. The program has been thoroughly tested from a user standpoint and a modification is being implemented daily in the hospital to provide a basic schedule. Given the basic daily schedule, studies requested after the start of the day may be incorporated on a best-fit basis or the program may be re-run.

Additional investigation may lead to a procedure which would guarantee an optimum schedule. An alternative would be to statistically validate the model with a much larger set of actual study schedules.

REFERENCES

1. CHEN HC & RICHARDS JE (1974) A description of the selection and testing of a set of random number seeds. *Proc. of the Twelfth Annual Southeast Region Conference–ACM.*
2. CONWAY RW, MAXWELL W & MILLER LW (1967) *Theory of Scheduling.* Addison-Wesley, Reading, MA.
3. EMSHOFF JR & SISSON RL (1970) *Design and Use of Computer Simulation Models.* Macmillan, New York.
4. JOHANNES JD (1975) A computer aided nuclear medicine patient scheduling and reporting system. Unpublished *MSc Thesis,* The University of Alabama in Huntsville, USA.
5. PATTON DD, STAAB EV & BRILL AB (1972) *Guide to Nuclear Medicine.* Vanderbilt University Hospital, Nashville, TN.

ADDRESS FOR CORRESPONDENCE: *Dr JD Johannes, Associate Professor of Computer Science, The University of Alabama in Huntsville, Huntsville, Alabama* 35807, *USA.*

Vol. 6
1978

OMEGA. The Int. Jl of Mgmt Sci. Vol. 6. No. 6. pp. 531–539
© Pergamon Press Ltd 1978. Printed in Great Britain
0306-0483/78/1201-0531$02.00/0

A Management Science Application in the Courtroom: A Jury Selection Model

CHARLES W GROSS

Illinois Institute of Technology, USA

HARVEY NUSSBAUM

Wayne State University, USA

(*Received June* 1978)

This paper reports the results of a study involving the application of management science techniques to the area of jury research. A public opinion survey was conducted, entered in evidence in a motion for change of venue, and the survey data used to develop a jury selection model used by the defense in a second-degree murder trial. The case had involved considerable pre-trial publicity. From a defense perspective, the defendant received the least undesirable guilty verdict. Management science techniques employed involved statistical sampling, linear discriminant function, and regression analysis. Suggestions for other applications in jury research are offered.

INTRODUCTION

MANY of us are aware of the important historical role played by management science practitioners in such legal matters as trademark and brand infringements suits, price discrimination cases, impaired earnings determination in personal injury suits, and a host of similar matters [3,4]. Within the last few years, the legal profession has increasingly recognized that the knowledge and analytic skills of management scientists is broad in scope, and has consequently turned to them for assistance in various legal matters. Perhaps one of the most dramatic of these new areas is the involvement of behaviorally oriented management scientists in highly political, socially sensitive trials [6,7]. This article summarizes one such trial and traces the role played by management scientists.

BACKGROUND[1]

Late one evening, on a hot Summer night in 1975, a white retail merchant heard "all too familiar" noises coming from the parking lot of his establishment. The establishment, located on the north side of a mid-western city near an area scarred by urban riots in the late 1960's, had similar problems in the past. Vandals had practiced their trade on parked cars and local businesses with an alarming frequency. The merchant decided to investigate, and discovered a black youth, in his late teens, attempting to break into a parked car, presumably owned by one of the merchant's patrons.

The merchant yelled "Stop," but to the shopkeeper's alarm, the youth began to run. After a few steps, the youth pulled a glimmering object from his pocket. The merchant fired a shot from a 0.25 caliber pistol which struck the youth in a vital spot. Later, investigation revealed that the shiny object was a knife, not a gun, as the merchant testified he thought at the time.

Heavy police presence, circumstances surrounding the incident, the area's high level of unemployment, possible racial tension, and a host of other factors may have all contributed to the situation. Whatever the reasons, a contained civil riot ensued. Stores were broken

[1] This court case remains under appeal. Consequently, certain details have been omitted so as to preserve confidentiality.

into, looting took place, fires were started, and a passing motorist was pulled from his automobile and beaten to death.

The merchant was arrested and charged with second degree murder. Perhaps in an attempt to quell the disturbance, many elected, appointed, and other civic leaders responded by stating publicly that the merchant should be punished because he was guilty of "triggering the riots." Several radicals displayed posters contending that the defendant was a "cold-blooded killer." The merchant's store was firebombed, and even the President of a university located in the general area was reported to state that the area was not safe after "the merchant started the riots." In sum, local public sentiment appeared very negatively biased.

TWO ISSUES

Attorneys for the defense were concerned that the defendant could not receive a fair trial, since it might not be possible to select an unbiased jury from residents within the city because of the case's extensive pre-trial publicity. Consequently, the authors were engaged to conduct a behavioral research study with two explicit objectives:

1. To ascertain the extent of pre-trial, deep-rooted notions of guilt or innocence among city residents eligible for jury selection. This was necessary for use in a defense motion for change of venue, i.e., to have the trial location changed to a city outside the actual locus of the crime.

2. To develop a scoring model (in the event of a denial of a motion for change of venue) to assist defense attorneys in selecting specific jurors to be included in the panel.

METHODOLOGY

Questionnaire

A telephone survey questionnaire was developed and pre-tested. In total, 29 questions were included. There were three categories of questions: demographic, direct case-related, and concealment of the study's intent. To obtain demographics, included were questions of race, sex, employment and occupation, age, marital status, and education. Fifteen questions were constructed to reveal awareness and attitude towards events directly related to the specific case, such as: (a) the shooting itself, (b) the ensuing riot, and (c) other case-related events. Included were two questions related to valence of attitude toward guilt or innocence and a measure of attitude conviction. These two questions, used to construct the dependent variable, are elaborated upon later in this article.

Finally, eight disguise questions related to other prominently known trials were asked to both conceal the true purpose of the study and to reveal possible clues as to juror voting behavior. Subjects were told only that a survey of news events was being conducted, not that its purpose was related to a specific case. For example, one question was (the survey was conducted shortly after an assassination attempt on President Ford):

People who try to assassinate the President should be executed. Do you:

_____ Agree strongly?
_____ Agree?
_____ Uncertain?
_____ Disagree?
_____ Disagree strongly?

The eight questions in this categoory were interspersed among the fifteen direct case-relatted questions.

Sample list

To select the sample, a random number generator was used to provide a list of random telephone numbers within the city [5]. The intent was to reach residential phones, including unlisted numbers (about 30% of the residential telephones in the city).

One concern was possible bias resulting from contacting subjects by telephone. Some persons eligible for jury duty undoubtedly had no telephones. From an operational perspective, due to time limitations, it was assumed that any such bias was immaterial. This assumption appeared realistic, as voters tend to be better educated, suggesting a higher proportion of phone users. Further, discussion with court officials indicated that the vast majority of persons called for jury duty do have telephones.

Vol. 6
1978

The survey

Jurors were to be drawn from voter registrations within the city. To assure that the sample represented the population, persons answering their phones were screened by first asking if they were registered to vote within the city. If they were not, they were asked if a city-registered voter lived in their household. If so, the registered voter was interviewed. If a business or institutional phone was reached or if no registered voter lived within the household, no other attempt was made to obtain an interview.

In order to reduce interviewer bias and to increase reliability, a professional survey firm was engaged to administer the questionnaire. Care was taken to assure that calls were made throughout a seven-day period at times ranging from early morning to early evening. Ten professional interviewers were used, comprising a mix of males, females, blacks, and whites. Two steps were taken, in addition to direct supervision, to validate the responses. First, chi-square tests were used to measure reponse consistency among interviewers; these were significantly related at the 95% level. Second, 10% of the respondents were called back to verify reported results.

In total, 806 telephone numbers were dialed. Of these, 236 (29.3%) were unanswered operating phones, 193 (13.9%) were to non-registered voters and hence purged, 64 (7.9%) reached persons who refused to participate, and 12 (1.5%) reached institutions, which were also purged. This yielded a final sample of 301 (37.4%) usable responses. Time and funding limitations—less than three weeks were available for the study, including pretesting and analysis—precluded all but a cursory attempt to call-back operating telephone numbers with no answers.

An exhaustive call-back design would have been preferred, but was not operationally feasible. However, the sample was validated by comparing the sample's demographics with population parameters using census data and election commission information, relating to, e.g. age, sex, occupation, employed/unemployed, race, etc.

Only voting proportions by sex were available from the City's Election Commission.[2] The sample's distribution of sex matched voter registrations almost exactly. Census data were used to validate the sample for other demographics. In all cases, except for education and marital status, the sample's demographics conformed to the parameters of the city's population for those old enough to vote. The sample did, however, have a higher education level than the general population. This was expected, nevertheless, as voting participation increases with education. An interesting finding related to marital status. A significant proportion of the sample indicated that they were "living together". Since census data did not report this information, subject marital status was not compared. Based on the above, it was concluded that the sample was reasonably representative of the population—a disproportionately large number of older persons, females, members of one race, or unemployed persons were not reached. After coding, a standard SPSS package was used to perform calculations required for all analyses.

RESULTS

Change of venue

In the jurisdiction in question, granting a change of venue required the presentation of evidence which indicated that a defendant was unable to receive a fair trial because of *deep-rooted* pre-trial publicity and cognitions of guilt or innocence. Recognizing the bias of merely presenting hand-selected witnesses testifying as to the extent of pre-trial prejudice (the historical procedure), the American Bar Association published standards which recommend the use of qualified public opinion surveys to support such motions [1].

To measure the extent of pre-trial cognitions of guilt or innocence, the following question was asked (administered only to those respondents familiar with the case's circumstances):

> "As you may know, the merchant is being tried for second degree murder. Now if you personally were to be on the merchant's jury, would you vote guilty, undecided, or not guilty?"

Table 1 presents the response.

Thus, 14.62 + 14.62 = 29.24% of the sample had pre-trial notions of guilt or innocence. To

[2] In the trial jurisdiction, demographic characteristics of registered voters other than sex were neither reported nor released.

TABLE 1. RESPONSES TO THE QUESTION, "HOW WOULD YOU VOTE"

	Number	Percent
Guilty	44	14.62
Undecided, or do not know enough about the case to vote*	213	70.76
Not Guilty	44	14.62
Total	301	100.0

* The total of 213 was made up of two parts; 62 respondents (20.6%) who were not familiar with the case plus 151 respondents (50.2%) who were undecided as to how they would vote. That is, it was assumed that if they were unfamiliar with the case, they did not have pre-trial cognitions of guilt or innocence (at least not at a level acceptable as evidence).

infer about the population, a standard error of proportions was calculated, as follows:

$$S_{\bar{p}} = \frac{P(100 - P)}{n} = \frac{70.76(100 - 70.76)}{301} = 2.62$$

Consequently, there was a 95% confidence that the true population level of pre-trial cognitions of guilt or innocence was $29.24 \pm 1.96(2.62) = 29.54 \pm 5.14\%$.

However, further modification was needed due to somewhat unique requirements of the jurisdiction. In lay terms, many states require only that there exist substantial pre-trial cognitions of guilt or innocence in order to demonstrate the need to grant a change of venue motion. The law in the jurisdiction of this case on the other hand, required that these feelings must be *deep-rooted* convictions which are unlikely to change. To assess the extent of deep-seatedness of these cognitions, the following question was asked of those who replied either guilty or innocent to the first question (the form of the question printed here is for those who said they would vote not guilty):

"You probably know how hard prosecutors work to convince you to change your mind. How sure are you about this case? Would you:

_____ Not change your mind about voting not guilty
_____ Possibly change your mind about voting not guilty
_____ Definitely change your mind about voting not guilty

Table 2 presents the results.

Those responding not guilty to the first question were asked the same basic question, except the term "defense attorneys" was substituted for "prosecutors" and "guilty" was substituted for "not guilty" in stating each of the possible responses. Thus, there were $8.97 + 8.64 = 17.61\%$ of the sample who were considered as having deep-rooted pre-trial cognitions about guilt or innocence. To infer about the population:

$$S_{\bar{p}} = \frac{82.39(100 - 82.39)}{301} = 2.20$$

Consequently, there was a 95% confidence that the true population level of deep-rooted pre-trial conviction was $17.61 \pm 1.96(2.20) = 17.61 \pm 4.31\%$.

Change of venue ruling

These data were submitted in testimony during a pre-trial hearing on the change of venue motion before the trial judge. After two and

TABLE 2. RESULTS MODIFIED TO REFLECT INTENSITY OF OPINION

Initial response	Initial number	Possibly or definitely change	Remaining	Percent
Guilty	44	(17)	27	8.97
Undecided or don't know*	213	35	248	82.39
Not Guilty	44	(18)	26	8.64
Total	301	0	301	100.0

* Those who stated that they might or would change their vote were added to this response category.

TABLE 3. DEPENDENT VARIABLE RE-CLASSIFICATION FOR SCORING SYSTEM (FROM TABLES 1 AND 2)

Code	Cognition	Number of responses	Percentage
1	Guilty, deep-rooted conviction (those unlikely to change)	27	9.0
2	Guilty, moderate conviction (those who said guilty, but might change)	17	5.6
3	Undecided (those who did not know of case or stated undecided)	213	70.8
4	Not guilty, moderate conviction (those who said not guilty, but might change)	18	6.0
5	Not guilty, deep-rooted conviction (those unlikely to change)	26	8.6
	Total	301	100.0

one-half days of testimony and counter-testimony, the judge ruled that the study was valid and a true reflection of the population. However, the change of venue motion was denied on procedural grounds. The judge indicated that the defendant could receive a fair trial within the city, since the *voir dire*[3] process in his opinion would be sufficient to purge any jurors with preconceived notions.

Scoring system

The second phase of the study, the development of a scoring system, began upon learning of the judge's decision. For the scoring system, the two questions relating to respondent cognitions of guilt or innocence were merged into one dependent variable, as presented in Table 3.

Next, cross-tabulations were prepared for each of the independent variables against the dependent so as to gain further insight into the meaning of the data. Some cross-tabulations were also prepared for the dependent variable against two or three presumed key independent variables (primarily demographic variables). Afterward, all variables were standardized using standard statistical procedures so that unit measures of all variables were compatible.

After standaridization, both linear discriminant and regression analyses were performed. Discriminant analysis was used five times, once for each dichotomous relationship of the dependent variable defined as one specific response category against all other categories. Next, multiple linear regression was performed. Comparison of the results revealed that the discriminant analysis did not materially improve the predictive power of the model over regression. Since the model was to be used in the courtroom by defense attorneys, simplicity was essential (only a pocket calculator was feasible in the courtroom). Consequently, an additive model was needed—one to enable prediction of all response categories. Therefore, all further analysis was based on multiple linear regression.

For developing the predictive model, the sample was randomly split into two groups: two-thirds of the subjects were used to build the model and one-third were used for cross-validation [2]. A multiple regression analysis was performed regressing twenty independent variables (disguise questions were excluded) against the dependent variable. Together, the twenty variables accounted for 56.8% (adjusted) of the response variance in the dependent variable (adjusted $r^2 = 0.56752$). Table 4 presents these results.

The equation presented in Table 4 was then used for cross-validation of the remaining one-third of the sample. Subject responses to the twenty questions were used to predict each of the subject's dependent variable scores. Predicted dependent variable responses correlated with actual responses at $r^2 = 0.5286$. After being contrasted with the model's adjusted $r^2 = 0.568$, it was concluded that the model was valid and could be used for prediction [2].

[3] Or pre-trial juror examination.

TABLE 4. PREDICTIVE EQUATION

Multiple Statistics	Analysis of variance	DF	Sum of squares	Mean square	F
Multiple $R = 0.79414$ R square = 0.63065 Adjusted R square = 0.56752 Standard error = 0.65755	Regression	20	86.37778	4.31889	9.8880
	Residual	117	60.58766	0.43237	

	Variables in the Equation				
	Variable identity	Beta	R Square	Standard error	F
	Age	0.122	0.066	0.068	3.204
	Race	0.066	0.114	0.067	0.987
	Education	−0.067	0.148	0.063	1.124
	Occupation	−0.045	0.172	0.063	0.500
	Marital Status	−0.080	0.172	0.066	1.480
	Sex	−0.043	0.497	0.062	0.472
	Question 12*	0.081	0.163	0.061	1.743
	Question 18*	0.510	0.493	0.070	53.6137
	Question 21*	0.221	0.557	0.072	9.398
	Question 2*	0.161	0.578	0.060	7.343
	Question 22*	0.127	0.597	0.067	3.619
	Question 20*	−0.064	0.600	0.062	1.065
	Question 19*	0.103	0.603	0.065	2.496
	Question 16*	−0.102	0.614	0.060	2.882
	Question 23*	−0.132	0.623	0.065	4.117
	Question 17*	0.066	0.626	0.062	1.142
	Question 9*	−0.055	0.628	0.061	0.822
	Question 6*	0.044	0.630	0.063	0.401
	Question 4*	0.028	0.630	0.061	0.209
	Question 7*	−0.027	0.631	0.064	0.185

* Due to the fact that the case is under appeal, and because of confidentiality, the defense has insisted that the actual identities of these questions are not to be released (see Footnote 1). In general, however, these variables refer to opinions about certain events only loosely related to the specific case, such as the ensuing riot, other news issues, and so forth.

That is, there was no significant shrinkage in the multiple r^2.

Scoring table

The next step consisted of preparing a scoring table to be used as a guideline for evaluating jurors during *voir dire*. Besides simplicity, (i.e. additive), a fundamental objective was to allow incorporation of as much information as could be collected about each potential juror. Therefore, standardized response scores for each independent variable were multiplied by the variable's beta coefficient (the estimated strength of the particular variable) calculated in the regression. To avoid working with decimals, all products were then multiplied by a constant of 10,000. The scoring table appears in Table 5.

To illustrate the system's use: If potential juror A's response to, say, question 18 was response category 2, response 4 to question 22, and response 2 to question 21 (and all other questions ignored for sake of illustration), the person's total score would be $-2{,}669 + 662 - 3{,}176 = -5183$.

Each potential juror's total score, after considering all independent variables, was then evaluated in relation to values of the standardized dependent variable—operationally termed "attitude," as presented in the bottom row of Table 5. The above illustrative person (score of −5183), therefore, would be predicted to be somewhere between an uncertain and a guilty—but might change his vote. This would be interpreted as a slight indication that a negative vote by the juror would be likely. Of course, an attempt was made to score responses to all significant questions, not just three questions as in the illustration. If the combined score of person B, for example, were 10,502, he or she would have been considered a very desirable juror. The greater the positive score, therefore, the greater was the considered likelihood that a juror would vote for innocence, and vice versa.

No attempt was made to consider actual

TABLE 5. SCORING TABLE

Variable statistics				Response weights—Scoring system							
Variable	Mean	Std. dev.	Obs	1	2	3	4	5	6	7	8
Age	2.82	1.25	150	−1771	−798	175	1148	2122			
Race	1.42	0.50	160	−571	722						
Education	3.18	1.24	159	1174	636	98	−440	−978	−1516		
Occupation	3.08	2.23	157	415	215	15	−185	−385	−585	−785	−985
Marital Status	2.39	0.81	165	1385	391	−602					
Sex	1.51	0.50	165	435	−419						
Question 12†	2.44	1.05	160	−1108	−341	427	1195	1962			
Question 18†	2.67	1.27	153	−6672	−2669	1334	5337	9340			
Question 21†	3.48	1.03	154	−5321	−3176	−1031	1114	3260			
Question 2†	1.88	0.33	166	4345	−595						
Question 22†	3.34	1.27	152	−2325	−1330	−334	662	1657			
Question 20†	2.40	1.21	154	732	208	−317	−841	−1365			
Question 19†	3.52	0.81	153	−3221	−1954	−687	580	1846			
Question 16†	3.29	1.38	160	1679	945	211	−523	−1257			
Question 23†	4.34	0.80	154	5491	3849	2207	565	−1077			
Question 17†	2.03	1.14	161	−602	−18	566	1149	1733			
Question 9†	2.55	0.92	163	936	333	−270	−873				
Question 6†	1.78	0.42	162	−743	212						
Question 4†	2.03	0.79	163	−365	−9	347	703				
Question 7†	1.28	0.45	166	169	−441						
Attitude	2.91	1.04	168	−18422	−8780	861	10502	20143			

* Adjusted for missing data.
† Confidential question identities, see Table 4.
‡ Dependent variable; response 1 is guilty—unlikely to change, response 2 is guilty but could change, response 3 is uncertain, response 4 is not guilty but could change, and response 5 is not guilty—unlikely to change.

probabilities of voting one way or the other so as to maintain simplicity. Neither was an attempt made to predict an exact category, as in the end only a guilty or not guilty verdict to a particular charge was possible. Instead, scores were used as a measure of relative positions along a guilty–not guilty attitude set continuum. However, attempting to predict exact probabilities appears worth considering in future applications. Then dynamic programming, perhaps, could be used as a decision tool for calculating optimal jury configurations.

Use of the scoring system

Several days prior to *voir dire*, lists of potential jurors were distributed to both prosecution and defense. As is often the practice in trial work, private investigators are used by the defense to obtain as much background information as possible about potential jurors. This information, along with court-required juror questionnaires, provided data for scoring potential jurors using the selection model.

In the case in question, the eventual trial jury consisted of fourteen jurors (twelve jurors plus two alternates who were randomly selected to be excused prior to deliberations). In the first phase of *voir dire*, a pool of 54 jurors was selected, all of whom passed for cause, i.e. they indicated that they held no biases or knowledge of principles in the case. Each person in this pool was also examined by the defense in order to elicit as much additional juror profile information as the judge would allow. Attorneys generally have some leeway in questioning potential jurors. In this case, only limited questions regarding the case were allowed by the trial judge, thus limiting the number of questions to be scored by the model.

In the second phase of *voirdire*, prosecution and defense alternately utilized their peremptory challenges until the final fourteen were chosen. Each side was allowed a total of twenty peremptory challenges. In addition to predicted dependent variable score, another factor utilized by the defense attorneys to select their preferred jurors were personal intuitions, i.e. "hunches" about a juror's potential hostility or support.

After reviewing the makeup of the jury pool, the defense decided to operate with a strategy of "minimizing its expected losses" in order to

TABLE 6. PREDICTED DEPENDENT VARIABLE SCORE FOR FINAL JURY

Juror	Predicted score*
1	2604
2	−988
3	−1052
4	649
5	−1398
6	−904
7	−591
8	919
9	−1366
10	−1358
11	681
12	1459
Average Predicted Score	−177

* Based upon utilization of six demographic variables and one attitudinal variable which could be collected during *voir dire* examination.

select the "best" of the final fourteen jurors from the pool which included many individuals, some of whom had net negative scores (likely to vote guilty).[4] The final twelve juror profile scores are presented in Table 6. The average predicted verdict score for these twelve jurors was −177, suggestive of an impartial jury (again, all standardized scores were multiplied by a constant of 10,000) but one possibly moderately inclined towards voting guilty. Any other jury selected from the available list of jurors would have predicted more confirmed levels of guilt. Thus, the jury finally selected minimized the defense's expected losses.

TRIAL RESULTS

The trial was relatively short, lasting about one week. The trial judge instructed the jury that there were four possible verdicts: guilty of second degree murder, guilty of manslaughter, guilty of reckless discharge of a firearm resulting in death, and not guilty. After two days of deliberation, the verdict reached was guilty of reckless discharge of a firearm, which is a high misdemeanor. While the verdict is clearly not as desirable as acquittal (from the defense perspective), and the case is currently being appealed, it was deemed a major (not necessarily total) victory considering the other possible guilty verdicts.

[4] Decision theory models of expectation were not used because of computation complexity.

Before concluding, two points need emphasizing. First, though such models are situation-specific and are not generalizable to other cases, i.e., the specific questions to ask, the importance of each question, and relative weights to be used for each response must be uniquely determined in each case, the same basic approach can be used in other cases.

Second, such a scoring system does not mean that trial evidence and circumstances are unimportant in influencing jury decisions. Clearly, convincing evidence and testimony is of major importance. But, when evidence is conflicting and unclear in establishing guilt or innocence, such a scoring system could be especially useful. A substantially different verdict could result from merely being able to objectively predict voting tendencies of a few jurors.

CONCLUSION

Both the recognition and role of the management scientist in legal matters has been broadened considerably in the last few years. This article presented the role of management scientists in a recent murder trial. The credibility of management scientists in the behavioral area withstood the challenge of legal counter-testimony by a noted behavioral scientist. Further, the analytical tools utilized by the researchers enabled the prediction of jury voting behavior. The model served as an aid to the defense in decision-making in an area where hunches and intuition have typically totally prevailed. Thus, the analogy to management scientists improving intuitive-based business decision-making can be made.

In most instances, the scoring system was heavily relied upon by the defense. But in two cases, jurors with negative scores were deliberately included because of specific defense attorney hunches. Of course, one can only speculate about what would have been the verdict if all jurors were evaluated exclusively with the scoring system. A more favorable verdict may have been possible, however.

The procedures outlined are not intended to serve as a guideline for future research in similar or other legal matters. Much has been learned from this and other legal cases in which the authors have been involved, especially in anticipating difficulties in a formal adversary relationship.

However, the application of management science research is sure to expand in other areas involving legal matters. Of course, there is no reason to assume similar jury analyses are limited to murder trials. Surely future research will become more intensely involved with assisting attorneys in jury selection and in other matters involving employment discrimination, liability claims, class action, malpractice, and other suits. Further, in complex cases where time and funding limitations permit, more sophisticated analytic models are likely to become employed, such as AID, MAID, and MCA. Nonetheless, the trend towards greater involvement by management scientists in legal matters is clear.

REFERENCES

1. American Bar Association Project on Minimum Standards for Criminal Justice (1968) *Fair Trial and Free Press*, p. 8.
2. DUNNETTE MD (1966) *Personnel Selection and Placement*. Wadsworth, Belmont, CA. pp. 240, 153–154.
3. EDEN P (1964) The use of economics and statistics in impaired earnings capacity cases. *The Practical Lawyer* **10**, 23–38.
4. KAIRYS D (1972) Jury selection: the law, a mathematician's method of analysis, and a case study. *Am. Criminal Law Rev.* **10**, 771–806.
5. SCHMIDT JW & TAYLOR RE (1970) *Simulation and Analysis of Industrial Systems*. Richard D Irwin, Homewood, IL.
6. SCHULMAN J, SHAVER P, COLMAN R, EMRICK B & CHRISTIE R (1973) Recipe for a jury. *Psychol. Today* **6**, 37–44.
7. ZEISEL H & DIAMOND SS (1976) The jury selection in the Mitchell–Stans conspiracy trial. *Am. Bar Foundation Res. Jl.* **1**, 151–174.

ADDRESS FOR CORRESPONDENCE: *Harvey Nussbaum Esq., Associate Professor, School of Business Administration, Wayne State University, Detroit, Michigan* 48202, *USA.*

BLANK
PAGE

Vol. 6
1978

OMEGA. The Int. J. of Mgmt Sci., Vol. 6, No. 6, pp. 541–544. Pergamon Press Ltd 1978. Printed in Great Britain

International Register of Theses and Dissertations

Boulay C (1976) *French: Inférence des traits de personnalité sur les critères évaluatifs: Application à un produit socialement visible.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Boule R (1976) *French: Le portefeuille international réel.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Hoffmann R W (1973/74) *German: Naturwissenchaftlich-technische Grossforschungsinstitute und naturwissenschaftlich-technische Intelligenz (Teil II).* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Jarousseau H (1976) *French: Modèle d'évaluation sur la marche des fusions de type conglomérat.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Lewin K (1971/72) *German: Die Entwicklung der Sozialwissenschaften in Göttingen im Zeitalter der Aufklärung 1734 bis 1812—Zur gegenseitigen Bedingtheit sozio-ökonomischer Prozesse und wissenschaftlicher Erkenntnis.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staaats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Linneweh K (1971/72) *German: Die Beurteilungsproblematik neurotischer Störungen im System der Sozialen Sicherheit.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Mpetsi C (1976) *French: Implications des fractionnements d'actions et motivations des compagnies.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Mutschler R (1973/74) *German: Bildungschancen im Hochschulbereich. Ein regionalanalytischer Vergleich zwischen wissenschaftlichen Hochschulen und Fachhochschuleinrichtungen in Norddeutschland.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Ohmstedt H (1973/74) *German: Die bankbetriebliche Anwendung der Metrapotentialmethode bei Emissionsgeschäften.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Quimet D (1975) *French: Modèle de planification des ressources humaines dans le système PPBE.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Philipp H-J (1973/74) *German: Sozialwissenschaftliche Aspekte von landwirtschaftlichen Siedlungsprojekten in der Dritten Welt unter besonderer Berücksichtigung tunesischer Projekte.* PhD Thesis, Georg–August-Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Politzer JP (1976) *French: Evaluation de l'information au sein des systèmes intégrés de gestion: une approche sur la théorie du contrôle.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Rodewald B (1973/74) *German: Die Portfolio Selection Theorie als Entscheidungshilfe für die Wertpapierdepotplanung.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Schirmer K P (1973/74) *German: Die neueren Bestrebungen zu einer Reform des Volkswirtschaftlichen Studiums in der Bundesrepublik Deutschland.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Schroder W (1973/74) *German: Betriebsanalyse und steuerliche Betriebsprüfung.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Tangermann S (1973/74) *German: Gemeinsame Agrarpreispolitik und nationale Wirtschaftspolitik in der EWG.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

Teulat D (1976) *French: Quelques éléments de base pour un contrôle du système d'information.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

Thibault D (1976) *French: Modèle de classification d'actions.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

THROLL M (1971/72) *German: Soziologie und Hochschulraumplanung—Soziologische Uberlegungen zum Raumkonzept einer praktisch-experimentellen Hochschule.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

TOCQUER G (1976) *French: L'utilisation de modèle de type "expectancy value" pour la segmentation du marché.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

VEIT K-R (1973/74) *German: Unternehmensverträge und Eingliederung als aktienrechtliche Instrumente der Unternehmensverbindung.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

VON ENGELHARDT M (1973/74) *German: Naturwissenschaftlich-technische Grossforschungsinstitute und naturwissenschaftlich-technische Intelligenz (Teil I).* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

VON WARTENBERG L (1971/72) *German: Kooperation im Handwerk—dargestellt an Beispielen aus Niedersachsen.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

VORMSCHLAG E (1971/72) *German: Ausgewählte politische Frauenzeitschriften der SPD, der USPD, der KPD (1890–1933) und der NSDAP (1932–1945)—Untersuchungen zu Weltanschauungen und Leitbildern der Frauenpublizistik.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

VOYER P (1976) *French: Les systèmes d'information de gestion: un cadre d'analyse et de conception.* MSc Thesis, Université de Sherbrooke. (Available at: la Bibliothèque Générale, Université de Sherbrooke, Québec J1K 2R1, Canada)

WEDELL H (1971/72) *German: Grundfragen des Minderheitenschutzes in Aktiengesellschaften und deren Lösung im Aktiengesetz von 1965.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

WEISE P L (1971/72) *German: Theoretische Grunde der mikroökonomischen Verteilungstheorie.* PhD Thesis, Georg–August–Universität, Göttingen. (Available at: Niedersächsischen Staats- und Universitäts–Bibliothek, Prinzenstrasse 1, 3400 Göttingen, Germany)

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

MCGUIRE M G (1973) Audio-Visual Aids to Computer Education. SM Thesis, MIT, Cambridge, MA.

MCNEILL W M F (1973) The Eurobond Market—A Descriptive Approach. SM Thesis, MIT, Cambridge, MA.

MILBURY D A (1973) A Study of Acquisition Strategy in Small Single Area Businesses. SM Thesis, MIT, Cambridge, MA.

NAULAPAA P E (1973) Capital Budgeting Procedure in a Telephone Company. SM Thesis, MIT, Cambridge, MA.

NEAL W D (1973) Formulation of a Forecasting Model for the Abrasive Grain Industry. SM Thesis, MIT, Cambridge, MA.

NETTER J M (1973) Computer Aided Picture Transformation. SM Thesis, MIT, Cambridge, MA.

NG B W (1973) Trading American Exchange Stocks with Filter Rules. SM Thesis, MIT, Cambridge, MA.

OCAMPO-GAVIRIA A (1973) Design of a Distribution System. SM Thesis, MIT, Cambridge, MA.

PETERSON R B (1973) Internal Entrepreneurship in the Aerospace Industry. SM Thesis, MIT, Cambridge, MA.

PIENS J D M (1973) Potency and Limitation of the Eurobond Market as an Integrative Force on the International Money Markets. SM Thesis, MIT, Cambridge, MA.

POHL T R (1973) Obstacles and Incentives for German Direct Investments in the US. SM Thesis, MIT, Cambridge, MA.

POTATE J S (1973) A Study of the Middle Level Manager's Perception of his Pay. SM Thesis, MIT, Cambridge, MA.

PROCTOR J D (1973) The Attitude of Lending Officers in Boston, Cleveland and Pittsburgh Toward Financing New Businesses. SM Thesis, MIT, Cambridge, MA.

PYBURN P J (1973) A Quantitative Analysis of the Risks and the Rewards of Arbitrage in Merger Hedges. SM Thesis, MIT Cambridge, MA.

QUAYLE A J (1973) Responsibility Accounting and Management-by-Objectives: A Comparative Study. SM Thesis, MIT, Cambridge, MA.

ROBSON E S III. (1973) Change Strategies: A Survey and Analysis. SM Thesis, MIT, Cambridge, MA.

RASHFORD N J (1973) The Strategic Process in an Academic Medical Center. SM Thesis, MIT, Cambridge, MA.

RATH D (1973) The Joint Negotiating Process: Case Study of an Indo–US Venture. SM Thesis, MIT, Cambridge, MA.

REDWINE S T Jr (1973) The Implementation of a Management Information and Control System for the Sloan School of Management. SM Thesis, MIT, Cambridge, MA.

RENICK D W (1973) Consulting and Planned Change in a Government Agency: One Consultant's Perspective. SM Thesis, MIT, Cambridge, MA.

RESNICK S R (1973) An Analysis of Managerial Problems in Power Plant Licensing and an Approach to a Licensing Framework. SM Thesis, MIT, Cambridge, MA.

ROBSON E S III (1973) Change Strategies: A Survey and Analysis. SM Thesis, MIT, Cambridge, MA.

RUBIN D J (1973) Homics—A Realistic Approach to Electronic Processing for Hotel Management. SM Thesis, MIT, Cambridge, MA.

SALOOM J A (1973) Organized Labor and Adjustment Assistance. SM Thesis, MIT, Cambridge, MA.

SAMPER J P (1973) Management of the International Advertising Function. SM Thesis, MIT, Cambridge, MA.

SAYRE P R (1973) Operative Decision Processes in Electric Utility Firms: A Case Study. SM Thesis, MIT, Cambridge, MA.

SCHERF J A (1973) Computer and Data Security: A Comprehensive Annotated Bibliography. SM Thesis, MIT, Cambridge, MA.

SCHNABL M E (1973) Software Packages and Planning Systems. SM Thesis, MIT, Cambridge, MA.

SCHONBEIN W R (1973) A Strategy to Study the Use of Radiology as an Information System in Patient Management. SM Thesis, MIT, Cambridge, MA.

SCHWEIKHARDT R G (1973) Labor Market Dynamics II. SM Thesis, MIT, Cambridge, MA.

SEAKAN R N (1973) Starting a New Manufacturing Firm: A Case Study. SM Thesis, MIT, Cambridge, MA.

SHOREY L J (1973) Evaluation of the Program Management System in Massachusetts. SM Thesis, MIT, Cambridge, MA.

SILVERMAN M (1973) A Study of the Financial Strategies that were Pursued by TWA and Delta During 1956–1971. SM Thesis, MIT, Cambridge, MA.

SMITH M G (1973) PCS: A Project Control System. SM Thesis, MIT, Cambridge, MA.

SPANGLER C W (1973) The Effects of Unanticipated Change in Dividends on Security Returns. SM Thesis, MIT, Cambridge, MA.

SPOON A (1973) An Interactive Planning and Control System for a Subscription Business. SM Thesis, MIT, Cambridge, MA.

STRASMORE M R (1973) The Strategic Planning Function Re-evaluated from the Organization Development Perspective. SM Thesis, MIT, Cambridge, MA.

TAMURA M (1973) Japanese Direct Investment in the United States: Opportunities for Future Growth. SM Thesis, MIT, Cambridge, MA.

TANG M Y-N (1972) Strategies for the Hong-Kong Textile Industry. SM Thesis, MIT, Cambridge, MA.

TANNENBAUM S L (1973) Tablecloth Corporation: A Study of Forecasting and Inventory Control in the Plastic Textile Industry. SM Thesis, MIT, Cambridge, MA.

TAVIS G G (1972) Model for Decision Making—Distribution of Federal Student Assistance Funds. SM Thesis, MIT, Cambridge, MA.

THOMSEN T R (1973) Regression Analysis for Forecasting and Controlling Expense Employees. SM Thesis, MIT, Cambridge, MA.

THORIN J P (1973) The Eurobond Market: Description and Impact on the International Long-Term Interest Rate. SM Thesis, MIT, Cambridge, MA.

THURBER R C Jr (1972) The Impact of an Interactive Risk Analysis Model on the New Venture Plans of a Group of Entrepreneurs. SM Thesis, MIT, Cambridge, MA.

TONEY E M (1972) Employment of Black Professionals in the Boston Labor Market. SM Thesis, MIT, Cambridge, MA.

TROUP S B (with VAN NIEL R W) (1972) Hospital Emergency Services: Modeling a Dynamic System. SM Thesis, MIT, Cambridge, MA.

TUNG K W (1973) A Computer Package for the Numerical Solution to a General Parabolic Equation in the Field of Finance. SM Thesis, MIT, Cambridge, MA.

TWINE E H (1973) A Dynamic Systems Analysis of Defensive Medicine. SM Thesis, MIT, Cambridge, MA.

URBANI D (1973) An Evaluation of Text Processing Systems. SM Thesis, MIT, Cambridge, MA.

VAUGHAN R R (1973) A System Dynamics Study of a Bank's Demand Deposit Accounting Operating System. SM Thesis, MIT, Cambridge, MA.

WADE O J (1973) A study of Organizational Climate and Managerial Motivation—Regulated versus Non-Regulated Companies. SM Thesis, MIT, Cambridge, MA.

WADE W M (1973) Health Care: Right, Consumer Good or Investment? SM Thesis, MIT, Cambridge, MA.

WAID P S (1973) Elasticity of Demand of Electricity Within Rate Structure. SM Thesis, MIT, Cambridge, MA.

WALL R E (1972) A Time Series Analysis of the Determinants of the Bid-Asked Price Spreads of United States Government Bonds. SM Thesis, MIT, Cambridge, MA.

WARNER E (1972) Mixed International Joint Ventures in the Exploration, Development and Production of Petroleum. SM Thesis, MIT, Cambridge, MA.

WASZKOWSKI H L III (1973) Financial Planning in a Small Company. SM Thesis, MIT, Cambridge, MA.

WATSON J P (1973) Marketing Planning for New Ventures in Mini-Computer Based Information Systems. SM Thesis, MIT, Cambridge, MA.

WEBB H G (1973) Collective Bargaining in the Railroad Industry: Progress and Problems of the 1960's Projected to the 1970's. SM Thesis, MIT, Cambridge, MA.

WEIR D R (1973) The Impact of Federal-Revenue Sharing on the Subsidized Family Planning Service Level in Massachusetts. SM Thesis, MIT, Cambridge, MA.

WESTERCAMP P J (1972) A Long Range Planning Model for New Product Introduction in the Computer Industry. SM Thesis, MIT, Cambridge, MA.

WHEELOCK K W (1972) Political Risk Insurance and US Investment in Spanish-Speaking South America: An Inquiry into the Rationale of Investment Insurance and the Climate for US Direct Investment. SM Thesis, MIT, Cambridge, MA.

WIGODSKY J (1973) An Interactive Planning and Control System for a Subscription Business. SM Thesis, MIT, Cambridge, MA.

WILLIAMS W D (1972) Medium and Long Term Financing of Capital Equipment Exports to the Developing Nations. SM Thesis, MIT, Cambridge, MA.

WILSON R E (1973) An Analysis of Negotiated Contracts in the Building Industry. SM Thesis, MIT, Cambridge, MA.

WINCH J B (1972) Change in Corporate Strategy: A Clinical Study. SM Thesis, MIT, Cambridge, MA.

WIREDU J K A (1972) Integrated Airline Planning and Control. SM Thesis, MIT, Cambridge, MA.

WOLFE P J (1973) A System Dynamics Study of a Bank's Demand Deposit Accounting Operating System. SM Thesis, MIT, Cambridge, MA.

WOOLRIDGE D (1973) Management Information Systems and Aerospace Program Management: A Utility Analysis. SM Thesis, MIT, Cambridge.

WRIGHT B R (1973) Automobile Leasing vs Ownership: A Staff Study. SM Thesis, MIT, Cambridge, MA.

WYNNE R G (1973) Change Strategies: A Survey and Analysis. SM Thesis, MIT, Cambridge, MA.

YAP C M F (1973) Development Banking Model—A Framework for Analysis. SM Thesis, MIT, Cambridge, MA.

YATES M J (1973) The Logic of "Flag Oil". SM Thesis, MIT, Cambridge, MA.

YOUNG A T (1972) Management Techniques for NASA Projects Involving Extramural Scientists. SM Thesis, MIT, Cambridge, MA.

ZIEN H B (1973) A Pilot Study of Planning and Control Systems Used in the Construction Industry. SM Thesis, MIT, Cambridge, MA.

ZOELLER G (1973) A Portfolio Approach to the Analysis of a Financial Intermediary—Case Study of the Federal National Mortgage Association. SM Thesis, MIT, Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.